Essays in Political Economy and Development

Dissertation
submitted to the Faculty of Economics,
Business Administration and Information Technology
of the University of Zurich

to obtain the degree of
Doctor of Philosophy
in Economics

presented by

Yikai Wang
from China

approved in July 2014 at the request of
Prof. Dr. Fabrizio Zilibotti
Prof. Dr. Mathias Hoffmann

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, July 16, 2014

Chairman of the Doctoral Board: Prof. Dr. Josef Zweimüller

Acknowledgement

In the past six years of my PhD study, I have received valuable support from many people, including colleagues, family and friends.

I am especially indebted to my advisor, Prof. Fabrizio Zilibotti. Benefiting so much from his generous support, I have gained valuable knowledge, not only in Economics, but also on how one can help others and can be a positive influence on others in academia and in life. I am very thankful to the faculty and the colleagues in the Department of Economics at the University of Zurich, who have been extremely helpful and inspiring during my PhD years. I am also extremely grateful to Prof. Daron Acemoglu. Without his invitation, I wouldn't have been able to spend an exciting year at MIT, which opened a whole new world to me and helped me know about and get interested in frontier research.

It was my great pleasure and fun to work with, be inspired by, and learn from my coauthors, Simon Alder, Zheng Song, and Kjetil Storesletten.

Last but not the least, I am very thankful to my parents, Xinlong Wang and Yun Chen, and my wife Wenrong, for their great patience and endless support. Without their support, I wouldn't have the privilege to enjoy research and pursue an exciting life in academia.

Overview

The focus of this dissertation is political and economic developments in emerging economies. Political systems shape economic institutions and therefore are crucial for economic development, especially in the developing countries, where political institutions vary a lot and differ greatly from the developed countries. Moreover, their political systems can change rapidly and go back and force. For these reasons, studying their political developments is crucial for understanding their economic development and obtaining more precise predictions on their future paths. This dissertation studies economic and political developments of emerging economies such as China, and aims to develop better understandings on political and economic phenomena in developing countries such as ethnic polarization, to think about their future developments e.g. China's growth potential, and to provide suggestions on appropriate economic policies such as pension reforms.

In the first chapter, "Will China Escape the Middle-income Trap? A Politico-economic Theory of Growth and State Capitalism", I study the following questions: is China's rapid growth sustainable if the labor and capital market distortions persist? Will democratization occur given that Chinese middle-class are supportive of the regime? To answer the above questions, this paper proposes a politico-economic theory, as follows. In oligarchy, a political elite extracts surplus from the state sector and taxes the private sector, but it also needs political support from sufficiently many citizens to maintain its power. "Divide-and-rule" strategy is implemented to guarantee such support: state workers receive high wages and become supporters of the elite, while wages of private workers are reduced due to the policy distortion. In the short-run, the low wages in the private sector lead to rapid growth of the private firms and total output. However, long-run growth is harmed by capital market distortions favoring the state firms. The theory suggests that the economy develops

along a three-stage transition: "*rapid growth*", "*state capitalism*", and two cases in the third stage: "*middle-income trap*" or "*sustained growth*", depending on whether democratization occurs endogenously. The theory is consistent with salient aspects of China's recent development and gives predictions on China's future development path.

The second chapter "Sharing High Growth Across Generations: Pensions and Demographic Transition in China" is a joint work with Zheng Song, Kjetil Storesletten and Fabrizio Zilibotti. This chapter studies the appropriate pension system for fast growing economies. We focus on China as an important example. More specifically, we build a overlapping generations model with endogenous labor supply to analyze intergenerational redistribution in emerging economies, where growth is high in the beginning but declines gradually. We calibrate the model to China, and analyze different pension reforms and their welfare implications. The current system is not financially sustainable, and a reform to reduce the replacement rates is necessary. However, it is not necessary to reform now. Delaying the reduction of replacement rates brings large welfare gains for the current generations who are poorer than the future generations whose losses are small. We also find that fully funded system is not appropriate for fast-growing countries with financial frictions, because it harms current generations, while bringing only small gains to future generations.

Simon Alder is the coauthor of the third chapter "Divide and Rule: An Origin of Polarization and Ethnic Conflict". In this chapter, we propose a theory of ethnic conflict to study its origin, dynamic evolution and consequences. If a society is composed of two ethnic groups and is under control of a political elite, the elite may initiate conflicts between the people of the two groups strategically to polarize the society and thus sustain its ruling position. We model polarization as a lack of trust, which is shaped by economic interactions between different groups. Higher trust implies larger expected gains from trade, and motivates the people of two ethnic groups to trade, interact with each other, and even cooperate to initial a revolution against the elite. Therefore, the elite prefers to keep trust level low by starting a conflict and thus interrupting trade, whenever it faces a large threat of revolution. Polarization and ethnic conflicts are therefore endogenous consequence of political struggles. This theory can also be extended to understand nationalism and conflicts between two countries.

Contents

List of Tables

List of Figures

Chapter 1

Will China Escape the Middle-income Trap? A Politico-economic Theory of Growth and State Capitalism

1.1 Introduction

China has by now been growing at a stellar rate for over 3 decades. While this is generally acknowledged to be a great historical achievement, there is major controversy on how far in time and scope the Chinese success story can go. The optimists argue that China can provide a new model for growth as an alternative to the liberal democracy growth model - the Washington Consensus. For example, in a debate hosted by The Economist (see also Musacchio and Lazzarini (2012)), Aldo Musacchio argues that China's hybrid form of capitalism can become a new growth model for the 21st century. In his view, such a model offers three very attractive features: less pronounced recessions, focus on long-term investing and producing world champions. These considerations make him optimistic about the sustainability of China's future growth, and even about the possibility that China can become a role model for other developing and emerging countries. In contrast, critics predict that the growth rate will soon slow down. For example, Acemoglu and Robinson (2012) argue that China's extractive political institution is not compatible with innovation and high growth in the long run. In their view, although the growth process driven by

catch-up, import of foreign technology, and export of low-end manufacturing products may continue for a while, it is deemed to come to a halt as soon as China reaches the living standards of a middle-income country.

The pessimistic perspective of Acemoglu and Robinson raises a number of questions. If growth under the current regime slows down, as they predict, will this trigger changes in the political system? Will unsatisfied citizens oust the oligarchy and allow growth to be resumed under a more democratic system? Or, alternatively, will the oligarchy be able to retain sufficient support even in a low-growth economy? On the one hand, modernization theory suggests that the first scenario is likely to occur. But, then, one can argue that it may have been right for China to adopt its hybrid form of state capitalism to achieve high economic growth in the catch-up stage, and then switch to liberal democracy when state capitalism runs out of steam. The Chinese model, in other words, could be a model of transition, albeit not of long run growth for mature economies. On the other hand, this view may well be overly optimistic: at the time of transition, the political elite may be unwilling to give up state capitalism and stick to power in order to keep control on political power and economic resources, as we see in countries like Venezuela. In the language of Acemoglu et al. (2006a), state capitalism may be appropriate to promote growth at an early stage of development, but may become impossible to reform at a later stage when it becomes a burden on further economic development.

To answer these two questions about China's growth and political transition, this paper proposes a theory of politico-economic transition. A two-sector dynamic general-equilibrium model is built and calibrated to China's economy. Moreover, the theory is consistent with salient aspects of China's recent developments, including: rapid growth with low wages and large state investments. Most notably, the theory can explain the high support of the Chinese middle-class to the political regime.

In this theory, a political elite is able to extract surplus from state firms and tax the private sector, however, it faces a political constraint, that is, support from sufficiently many citizens. To gain the support, it implements the "divide-and-rule" strategy. It creates a dual labor market, in which state workers receive high wages and private workers' wages are reduced due to the policy distortion. The state workers who benefit from the policy become the elite's supporters. Furthermore, to satisfy the political constraint, the elite finds

it optimal to distort the allocation of capital between state and private sector. The private sector contributes taxes, but also competes for labor with the state sector. So the elite first encourages the growth of the private sector but then restricts it when the growth of private employment turns into a threat for the elite's supporter base, i.e., state employees. Therefore, government policy and economic growth, follow different patterns in different stages of development. The economy develops along a three-stage transition, as follows. The first stage is "*rapid growth*", during which the GDP share of the private sector grows fast, triggering high reallocation and productivity growth. In this stage, private firms benefit from the distorted low wage in the private sector induced by the policy. The government supports privatization as this increases its tax revenue. However, as privatization goes on and the state employment share declines to a critical level, the economy enters the second stage - "*state capitalism*". In this stage, the elite over-invests in the state sector to keep the state employment sufficiently high. The government also imposes gradually increasing financial repression to limit the growth of private firms. Growth continues to be high due to large state investment, but the financial repression on private firms causes a slowdown. As the private sector capital keeps growing (largely through self finance), two possible outcomes emerge. The first is the "*middle-income trap*": the state over-investment and financial repression on private firms continue, but due to decreasing return to capital and the capital market distortion, the efficiency loss grows larger. Finally, the growth stops before the output converges to the level in democracy. This happens when the cost of retaining the state sector is low, i.e., when the number of supporters needed is small. The other possible outcome is democratization leading to "*sustained growth*". In this case, the elite finds it too costly to keep investing in the less efficient state sector and therefore chooses to democratize. State over-investment and financial repression on private firms both disappear and the economy keeps growing in democracy.

The first two stages in the theory are consistent the recent development in China. First, the distorted low private sector wage helps private firms and the economy grow rapidly. Between 1995 and 2007, the private employment share increases from 40% to 80% (see more details in section 1.2). This era of fast privatization implies large efficiency gain and "rapid growth", as the first stage of the theory. However, afterwards, private sector employment share stops growing. Private firms face tighter financial constraints while around 60%

of investment and the majority of bank loans are diverted to less productive state firms.[1][2] This capital market mis-allocation in favor of state firms implies that the economy is entering the "state capitalism" stage. Second, the middle-class, consist largely of state sector workers and private entrepreneurs, are the beneficiaries and supporters of the regime. This is because state workers receive high wages, and entrepreneurs benefit from the cheap and abundant labor in the private sector . Chen and Lu (2011) and Tsai (2007) document that the Chinese middle-class, including state employee and private entrepreneurs are "achieving their material interests without pursuing any real freedom". This phenomenon will be discussed in great details in section 1.2. Besides the above phenomena, the theory is also useful to understand a few more, including: high capital labor ratio in the state sector, low and decreasing state sector capital return, high and non-decreasing private sector capital return, etc..

The third stage of the transition in the theory provides an answer to the questions on China's future political and economic developments. The model, calibrated to China's economy, predicts that the economy will enter the "middle-income trap". The reason is the relatively low cost of retaining enough supporters in the state sector. One the one hand, the government is economically powerful and it is able to invest and maintain a large state sector, because it controls the banking sector and holds abundant financial asset, including the huge foreign reserve. On the other hand, the current government is politically powerful, meaning that unless a very large fraction of citizens are against it, it can retain its control over the country. In other words, a relatively small fraction of supporters is sufficient to maintain the regime and state employment share doesn't have to be too high. Given these conditions, the current regime and policy distortions will persist, which will eventually slow down the growth before China becomes a rich country.

Is China doomed to fall into the middle-income trap? Is there any way to change the conditions and move China to the other development path - "sustained growth"? In principle, political and economic reforms can change the conditions which determine the development path. We model reforms as a bargain between the elite who care about their

[1] 80% of bank loans are received by large firms who contribute to only 30% of GDP and 20% of employment. Most of them are state firms.

[2] Hsieh and Klenow (2009a) estimate that the total factor productivity (TFP) of state firms is 42% lower than the TFP of domestic private firms.

own interest and technocrats who care about the productivity. Many reforms, such as such as financial reforms, political reforms and state firm reforms, are suggested by researchers and China watchers. But will the elite agree to implement those reforms which may harm its interest? With the aid of the model, we evaluate how different reforms affect growth and also elite interest. This helps us to think on which reforms will face strong resistance from the elite and which are more likely to be implemented by the government.

Our theory is related to three strands of literature. The first one is on China's economic growth with resource re-allocation from the state sector to the private sector. Song et al. (2011a) construct a two-sector growth model to study how the capital and labor reallocation from the state to the private sector leads to economic growth. Brandt and Zhu (2000, 2010) document the contribution of private firms to growth and show that the government tries to maintain the state sector employment share. In these studies, the state sector employment share is either determined by the pure economic force or assumed to be exogenously set by the government. The government's political constraint introduced in our model provides the micro-foundation for the endogenous evolution of the state sector employment share, which gives a better match to the data and a richer prediction on the further growth trend.

Second, our theory contributes to the study of "middle-income trap", i.e., the significant slowdown of economic growth when a country's GDP per capita reaches the middle level. See Gill and Kharas (2007) and Eichengreen et al. (2013). Fatás and Mihov (2009) argue that this is because growth from low income to middle income doesn't require good institutions but only right policies, but good institutions are necessary for high income countries. Without improved institutions, rapid growing countries will "hit the wall". This discussion is also heated in the public but there are in lack of theoretical frameworks to provide guideline for it. With the model, we formally show how a country can grow within the extractive institution but the growth stops at a middle level if there's no reform on the political institution. Also, we study how a country can avoid falling into the middle-income trap.

The third is the literature on the relation between the political and economic institutions. Acemoglu and Robinson (2012) focus how the political institution affects economic performance in the long run. The argue that the *extractive* political institution in a non-democratic country is detrimental to economic growth. On the other hand, Lipset (1959) and Fukuyama (1992) study how the long-term effect of economic development on the

political development. Their modernization theory argues that the economic development will ultimately leads to political modernization, i.e., liberal democracy. Our contribution to them is that in addition to the effect of one on the other, we emphasize the inter-dependency of the economic institution and the political institution, to give a more complete picture both in the short- and the long-run.

The rest of the paper is organized as follows. Section 1.2 shows some important empirical facts on China's political-economic development. Section 1.3 discusses a two-sector dynamic growth model with the three-stage political-economic transition. The first two stages explain puzzles in China's recent development, while the third stage predicts future politico-economic trend. Section 1.4 concludes.

1.2 Empirical Facts on China's Recent Development

In this section, we discuss the key phenomena and puzzles in China's recent development, including: (1) large wage gap between the state and the private sector; (2) the middle-class's low support for democracy, as the opposite of the conventional wisdom; (3) the partial privatization; and (4) financial repression on private firms.

1.2.1 Large State-Private Wage Gap

China's rapid growth is accompanied by increasing inequality: the Gini index grows from 0.36 in 1992 to 0.47 in 2010 as in the official report but as high as 0.6 in various survey data. One important contributor to the inequality is the increasing state-private worker income gap.State workers enjoy a wage premium of around 20% to 30%, all the characteristics - age, education, industry, region and so on - being equal, as documented by Ge and Yang (2012) with the Urban Household Survey 1992-2007. Their result is reproduced in figure 1.1. In contrast, the wage premia of state workers in Canada, Germany and the U.S. are estimated to be lower than 5% after the 90s. See Melly (2002), Mueller (1998) and Poterba and Rueben (1994).

On the one hand, relatively low private sector wage help private firms to grow, and also contribute to China's growing export and output. Meanwhile, because of the high state

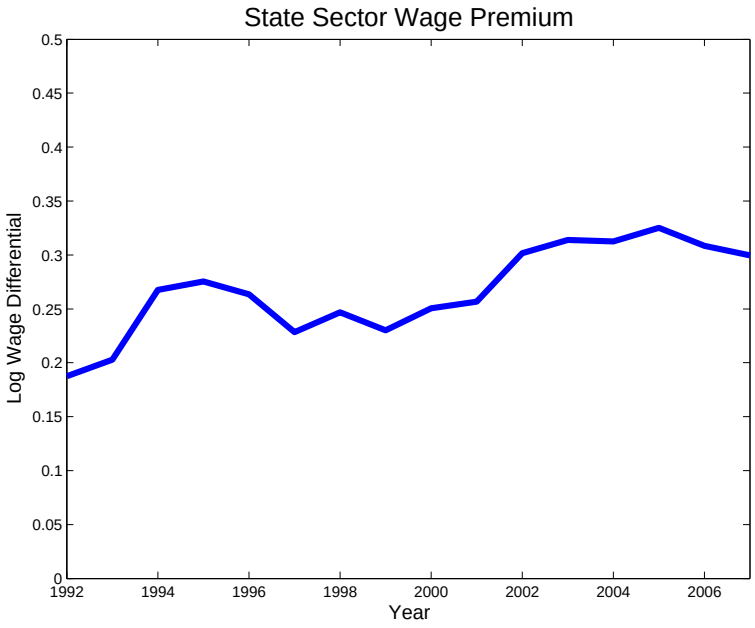

Figure 1.1: State sector wage premium.

sector wage, state sector jobs are very popular and in short supply in the market, especially for the newly graduated - on average 169 applicants for 1 position in 2013 Shanghai. In principle, state firms can reduce wages, hire more workers and enjoy higher profits. Why doesn't this happen? Why is there a large and persistent wage gap between the state and private firms? This is one puzzle that our theory aims to explain.

1.2.2 The Middle-class's Political Support for the Regime

China's middle-class largely consist of state workers and private entrepreneurs. State sector workers are beneficiary and supporters of the regime, because of the high wages they receive. Private entrepreneurs benefit from the cheap and abundant labor in the private sector, and are also satisfied with the current policies. Tsai (2007) documents that the Chinese entrepreneurs are "achieving their material interests without pursing any real freedom", different from "the business classes in historical England, France and the United States" who "have risen up against the government to defend material interests". The Chinese middle-class are not supporters of democracy, on the contrary to the conventional wisdom that the

Support for Democracy	coefficient
Employment in State Sector	-1.23**
Middle-class Membership	-0.54**
Party Membership	-0.37

Table 1.1: Support for Democracy Among Different Groups.

middle-class are the driving force for democratization. This phenomenon is systematically documented by Chen and Lu (2011). They use a survey data of 2810 individuals, collected in three Chinese cities in late 2006 and 2007 to estimate how the individual's political opinions depend on his/her characteristics, especially the social group identity. They find that state sector employment and the middle-class membership are negatively correlated with the support for democratic values.[3] For example, only 24.9% of the middle class support multi-party competition, while 38.7% of the lower-class do. Only 22.9% of the middle class agree demonstration should be allowed, while this number is 35.6% for the lower class. Similar patterns apply for other questions related to democratic values and institutions. To formally show the difference between the middle class and the lower class, the authors combine answers to multiple questions into one index of support for democratic values and institutions using factor analysis.[4] Then they run a regression of this index on individual characteristics, including the middle-class membership and employment in state sector. The coefficients for middle-class membership and state employment are significantly negative, and they are even stronger predictors compared to the party membership, suggesting that the economic interest plays a more important role than ideology. In other words, the middle class, including many state sector workers, are more supportive for the current political system. Part of their regression results are reproduced in the table 1.1.

[3]The authors define class according to the employment status. Individuals with jobs which usually pay low wages are classified as the lower class, including blue-collar workers, unemployed and self-employed with very little capital. The middle class mainly consist of white-collar workers. The authors distinguish the middle class from private entrepreneurs, but report that private entrepreneurs hold similar political opinions as the middle class. So their findings on the middle class can be applied to private entrepreneurs.

[4]The survey data contain four dimensions of questions on support of democratic values, including right consciousness, valuation of political liberty, support for participatory norm and support for competitive election. The index for support for democratic values and institutions is the constructed as the single dominant factor using factor analysis.

1.2.3 Partial Privatization

Since the fifteenth national congress in 1997, the state firm reform has transformed state firms into independent units who are responsible for their own operations, decisions and profits. Unprofitable state firms bankrupt and exit the market while more private firms enter and replace them. The privatization was very rapid for a couple of years. As the blue line in figure 1.2 shows, the employment share of state sector in the urban area decline from 53% in 1997 to 28% in 2002, and 22% in 2006. After that, the privatization slows down and the state employment share stagnates at around 20%. If we focus only on the manufacturing, mining and construction, represented by the red line, the trend is similar though the state employment share stops declining at around 40%, and even slightly increases in 2011. The phenomena, called "the state advances as the private sector retreats", is intensively discussed in the public and becomes a major concern for China's growth. Moreover, the government seems to intentionally keep the state sector alive. For example, in the closing announcement of the Third Plenary Session of 18th Chinese Communist Party Central Committee, it is stated that "China will stick to the dominant role of public ownership, playing the leading role of the state-owned economy, while encouraging, supporting and guiding the non-public sector..." Why doesn't the privatization continue - as a pure economic model would predict - until all the inefficient state firms exit?

1.2.4 Financial Repression on Private Firms

The direct reason why the inefficient state firms are still alive is that they get cheap loans from state banks while private firms get much less loans though their capital return is higher. Song et al. (2011a) report that while state firms finance more than 30% of their investment through bank loans and government budget, this number is less than 10% for private firms. Their result is reproduced in figure 1.3. Huang (2008) argues that the disadvantage of the private firms in the financial market is due to government policies in favor of state firms and repressing private firms, and this capital market distortion is getting more severe over time. Brandt et al. (2013) estimate that the capital wedge, i.e. the ratio of costs per unit of capital between state and private firms has increased in all the provinces, on average from 4.2 in 1996 to 6.8 in 2007.

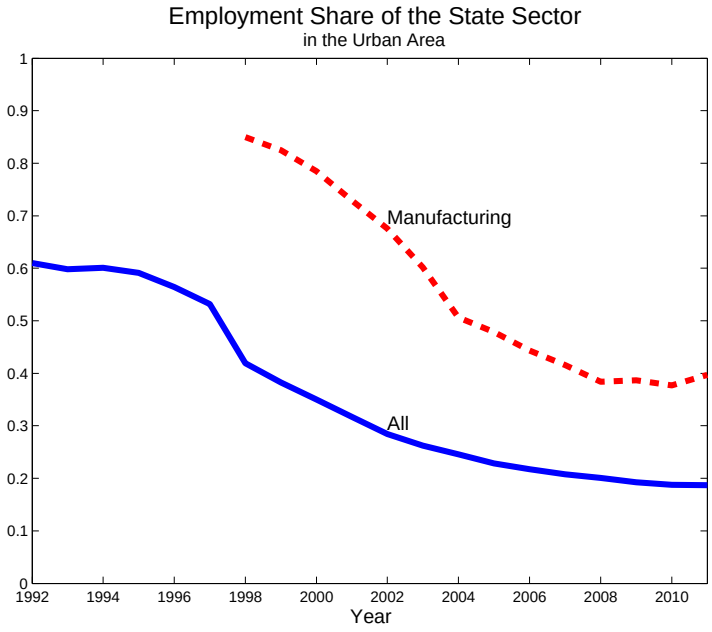

Figure 1.2: State sector's employment share.

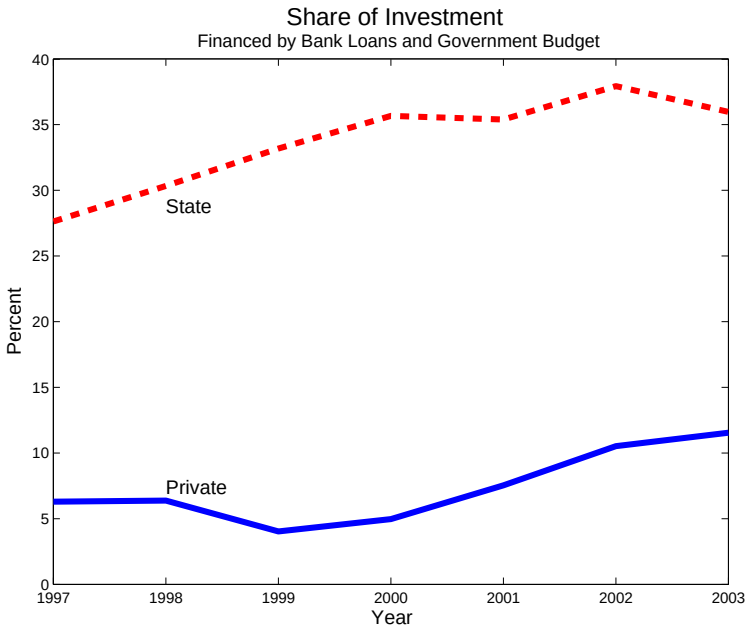

Figure 1.3: Share of investment financed by bank loans and government budget.

1.3 The Model

In this section, we present a theory of politico-economic transition to address the questions raised above on China's future economic growth. We build a two-sector dynamic general equilibrium growth model in which agents also make choices affecting the sustainability of the political system. We first discuss the general properties of the model and then its implications for China with the aid of a calibrated economy.

1.3.1 Environment

The model economy is populated by three classes of infinitely many agents: elites (e), private entrepreneurs (p), and workers (w). The population of workers is normalized to measure 1, and the population sizes of elites and private entrepreneurs are assumed to be very small and of measure 0.

There are two sectors and two types of firms. State (S) firms produce in the state (S) sector, while private (P) firms in the private (P) sector. There are infinitely many of them. They produce the same final goods with capital and labor to maximize profits. They are different in two aspects. First, ownership: S firms are owned by elites, while P firms by private entrepreneurs. Second, productivity: S firms are less productive than P firms. Technology of S and P firms are described by the following production functions:

$$\begin{aligned} Y_S &= (z_S K_S)^{\alpha} L_S^{1-\alpha}, \\ Y_P &= K_P^{\alpha} L_P^{1-\alpha}, \end{aligned}$$

where $z_S < 1$, K_S, K_P are S and P sector capital while L_S, L_P denote for S and P sector labor, respectively.

Elites provide capital to S firms while entrepreneurs to P firms. They earn income from the capital returns. They live for infinite periods, and are forward looking. Their instantaneous utility is assumed to be logarithmic and the discount factor is denoted by β. Workers provide 1 unit of labor inelastically. For simplicity, we assume that workers live hand-to-mouth, i.e. they consume all the income every period.

Elites have access to the deep pockets of banks. In other words, they can borrow from banks and set S sector capital without constraint. An entrepreneur finances P firm capital partly with her asset, and partly with bank loan. However, she faces the financial constraint: the bank loan can not exceed $\eta - 1$ fraction of her asset. In other words, the P firm leverage - ratio of capital over net asset - is bounded above by η. η is set by the government within a region: $\eta \in \left[\underline{\eta}, \bar{\eta}\right]$. Furthermore, we assume banks can borrow and lend in the international bond market at the interest rate r and they compete with each other, so the interest rates for loans to state and private firms are both r, and the interest rates for elite and entrepreneur savings in the bank are also r.

The above setting of the financial market is similar to Song et al. (2011a) while the difference is that this paper allows the financial constraint - P firm leverage η in our context - to be endogenously determined by the government. In China, state firms enjoy good connections with state banks and are backed by the government, so they can get bank loans relatively freely. In contrast, the private firms have limited access to external financing, due to the lack of state guarantee and connections with state banks, but also largely because of the government's policies that make state banks less willing to lend to private firms. The government can create barriers in loans to private firms, or directly give administrative instructions to banks (see Brandt and Zhu (2000)). $\underline{\eta}$ is the lower bound of the leverage. For example, $\underline{\eta}$ equals 1 if the strictest policy that the government can set is to order banks not to lend to private firms at all, but the private firms can still finance their investment using entrepreneurs' asset. $\bar{\eta}$ is the highest leverage if the government doesn't restrict the private firm financing at all. It is not infinity, because of the moral hazard problem, i.e., an entrepreneur with too much loans compared to her asset will choose to steal and run away.[5]

There are two types of political regimes: democracy and oligarchy. In democracy, the government is elected by majority vote; therefore, workers, given their dominating

[5]The logic is similar to Song et al. (2011a), as follows. Banks want to make sure that the borrower will not steal and run away. The borrower with asset s might be an entrepreneur who has access to production technology and obtain high return to asset $r_p > r$, or simply a worker who only saves all the money in other banks and gets return r. The borrower can choose to steal and run away, and a worker facing lower return than an entrepreneur, is more likely to do that. The bank has to guarantee that even a worker doesn't want to steal. Suppose that the bank can still take back $\tilde{\eta}(1+r)(l+s)$, where l is the loan. The borrower gets $(1-\tilde{\eta})(1+r)(l+s)$. So the incentive constraint for the worker as the borrower is $(1-\tilde{\eta})(1+r)(l+s) \leq (1+r)(l+s) - (1+r)l \Rightarrow l \leq \frac{\tilde{\eta}}{1-\tilde{\eta}}s$. Finally, we define $\bar{\eta} = 1 + \frac{\tilde{\eta}}{1-\tilde{\eta}}$, which is the maximal leverage.

population, determine the government policies. In oligarchy, elites control the government, but they still face the political constraint, i.e., support from a sufficiently large fraction of workers. Each worker, after being employed by a S or P firm and observing the government policies, decides to support the oligarchy or not based on the expectation on her income. The oligarchy is sustained if more than $\underline{L}$ fraction of workers choose to support it. If it gets less than $\underline{L}$ workers' support, democratization occurs. The setting on political constraint and political support in oligarchy is micro-founded on a sequential game between workers and elites. We leave the details of the sequential game in the appendix, but the logic that oligarchy needs political support from workers is quite intuitive and natural, as follows. Elites, given their small population, are not powerful enough to control the country alone, so support from some workers is necessary. On the other hand, elites also hold significant political power in oligarchy, so they don't need all workers' support but just a fraction of workers' support, which is $\underline{L}$.[6] Notice that $\underline{L}$ captures the relative political power of elites compared to workers. If elites are very powerful, they need only a small fraction of workers as supporters, in other words, $\underline{L}$ is small. If workers are well-organized and politically motivated, $\underline{L}$ is large.

In both political regimes, the government decides the tax rate and which groups to tax. We assume that tax payers can hide their income at the cost of $\bar{\tau}$ fraction of their income. This implies that if the tax rate is lower than $\bar{\tau}$, tax payers choose to pay the tax. Otherwise they hide the income and pay no tax. This is a simple way to model the exogenous tax rates, as used in Acemoglu (2008) and referred as "state capacity" in Besley and Persson (2009). The tax payers are different in democracy and oligarchy. In democracy, the government taxes elites and entrepreneurs. In oligarchy, the government taxes entrepreneurs and P workers, but not S workers. Basically, the government doesn't tax the ruling group and the necessary supporters of the ruling group. This is in fact optimal for the government. The

[6]The micro-foundation for $\underline{L}$ is the following: if elites and their supporters form a powerful enough coalition, oligarchy is sustained, as in Acemoglu et al. (2012). In their language, a coalition of agents is associated with certain level of political power, and if it is large enough, they decide political system. In oligarchy, elites as the ruling group are granted some level of political power, denoted as ω_e. Each worker has political power ω_w, and each entrepreneur has ω_p. The aggregate political power of entrepreneurs is 0 given its small size. Workers can change the political regime from oligarchy to democracy if and only if they form a coalition of size L_r whose power is larger than α, namely $\frac{\omega_w L_r}{\omega_w+\omega_e} > \alpha \Leftrightarrow L_r > \alpha\frac{\omega_w+\omega_e}{\omega_w}$, where α is exogenous. In other words, to sustain a oligarchy, there must be at least $1-\alpha\frac{\omega_w+\omega_e}{\omega_w}$ workers supporting it, and we can denote this size as $\underline{L}$.

government also decides other policies, including P sector financial constraint η, S sector capital K_S, S sector minimal wage w_S, and transfer. We will discuss them in greater details in the next subsection.

1.3.2 The Equilibrium Given Capital Allocation

The dynamic equilibrium consists of infinite periods, and each period can be decomposed into three stages: (1) determination of capital in S and P sectors, (2) the equilibrium of the labor market and political outcome in this period given capital allocation, and (3) decisions on consumption and saving. In this subsection, we first focus on the stage (2) of each period and study the equilibrium in that stage. It is crucial for understanding how the political outcome is determined by the government policies and economic power. We also don't consider how tax is determined in this subsection. So in this stage, the government can use S sector minimal wage to influence the labor market, firms hire and produce, and workers decide political support.

In democracy, workers do not want to impose distortive policies on the labor market or change the political system, because the competitive equilibrium in democracy maximizes the labor income. Moreover, the government, controlled by workers, collects tax from elites and entrepreneurs and transfers it to workers. In the competitive equilibrium. wages in S and P firms are the same and are equal to the marginal productivity of labor:

$$w^D = (1-\alpha)(z_S K_S)^{\alpha}\left(L_S^D\right)^{-\alpha} = (1-\alpha)(K_P)^{\alpha}\left(L_P^D\right)^{-\alpha}.$$

A worker's one-period income equals the wage plus the tax collected from entrepreneurs and elites:

$$\begin{aligned} y_w^D &= w^D + \tau^D\left(\pi_S^D + \pi_P^D\right) \\ &= \left(1+\tau^D\frac{\alpha}{1-\alpha}\right)w^D, \end{aligned}$$

where π_S^D and π_P^D are the capital incomes in S and P sectors, respectively and τ^D is the tax rate in democracy. The transfer to workers is $\tau^D\frac{\alpha}{1-\alpha}w^D$ simply because the tax base - capital income - is $\frac{\alpha}{1-\alpha}$ times labor income.

In oligarchy, the following events happen sequentially: (1) the government sets S sector minimal wage; (2) S and P firms hire workers; (3) S and P workers decide whether to support the current political system; (4) number of supporters determine the political outcome; (5) firm produce, labor and capital incomes are distributed; (6) the government collects tax and makes transfer.

First, the government chooses S sector minimal wage w_S to affect the labor market outcome and the economic benefits of S and P workers.[7] Without loss of generality, we assume that $w_S \geq w^D$ so the minimal wage constraint is tight.[8] Given the minimal wage, S firms employment is determined given the following assumption:

Assumption 1.1. *S firms maximize profits. They are free to determine the employment while they obey the minimal wage set by the government.*

So S firms choose labor demand L_S so that wage equals marginal productivity:

$$w_S = (1-\alpha)(z_S K_S)^{\alpha} L_S^{-\alpha}. \tag{1.1}$$

Furthermore, we make a second important assumption.

Assumption 1.2. *In oligarchy, the government can not make direct transfer to the ruled groups, including workers.*

So the final income of S workers is:

$$y_{wS} = w_S.$$

These two assumptions together imply the following: to increase S worker income, the government has to set a high S sector minimal wage, which distorts the labor market. We can see this in figure 1.4. Red and blue lines are the marginal productivities of labor in S and P sectors, respectively. The intersection of the two lines pin down the equilibrium in democracy: the S sector labor, wage and worker income in democracy are denoted as L_S^D, w^D and y_w^D. In oligarchy, w_S pins down S sector labor and its marginal productivity.

[7] Notice that we use w_S instead of w_S^O to simply the notation. We drop the superscript O for variables in oligarchy when there is no confusion.

[8] $w_S < w^D$ is equivalent to $w_S = w^D$.

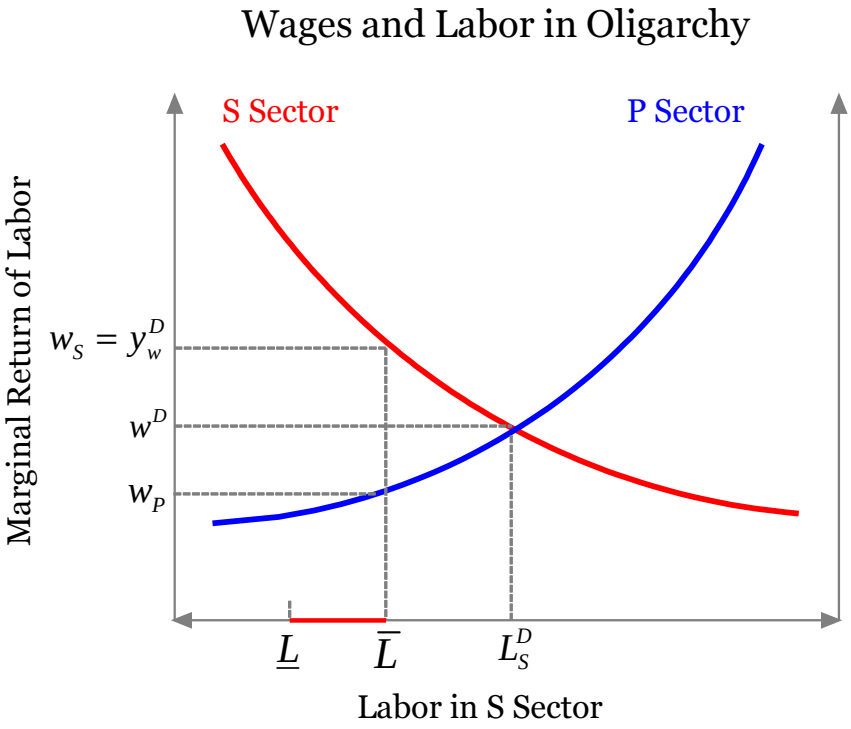

Figure 1.4: Labor allocation and marginal productivities.

The rest of labor is in the P sector and pins down the P sector wage w_P. We can see that setting w_S larger than y_w^D implies that the marginal productivity of S sector labor is higher than y_w^D and the S firms hire labor lower than $\bar{L}$. Observing the government policy on w_S, a S worker knows her one-period income in oligarchy. A worker supports oligarchy if and only if her income in oligarchy is higher than in democracy. This is very intuitive, hence we leave the discussion why this is optimal for a worker in the appendix. For simplicity, we assume that workers are myopic, so they care only about the current period income. So a S worker supports oligarchy if $w_S > y_w^D$.

Are the two assumptions and their implications reasonable? Why does the government have to use distortive policies to guarantee high incomes for S workers? If possible, the government may want to simply order S firms pay wage higher than the marginal productivity, or make direct government transfer to S workers. The two assumptions are reasonable for China, for the following reasons. First, China's 30 years of state firm reform is essentially delegation of rights from the government to firms so that they are incentivized to

maximize the profits. Nowadays, state firms are responsible for important decisions including hiring workers. Though the government can still affect the state sector wage through laws and regulations, for example, state firms must pay the pension tax for workers, buy the health insurance for workers, and so on, a state firm is free to decide labor to maximize its profit. Second, direct transfer from the government to workers is very rare. This more direct tool to provide economic benefits and solve political conflicts is difficult to implement in reality. This is not only the case in China, but is also discussed in other circumstances in the literature. One reason is the commitment problem. Acemoglu (2003) and Acemoglu and Robinson (2005) explain that though the state promises to make a transfer to the ruled group, the latter, without political power, gets no guarantee that they will eventually receive the transfer. So in many cases, transfer can not be used solve the political conflicts.The other reason is the high cost of government transfer due to local capture. This is supported by empirical evidences. Reinikka and Svensson (2004) document that 87% of the transfer from the central government to local schools in Uganda was not received during 1991-1995 due to local capture. This means that the cost of 1 dollar of transfer is as high as 7.7 dollars. For these two reasons, the government usually builds inefficient "white elephant" projects (see Robinson and Torvik (2004)) to guarantee the economic benefits for certain groups. In our model, state firms can be considered as a special type of "white elephants".

As we can see from figure 1.4, P sector wage is lower than the wage and worker income in democracy, due to the general equilibrium effect. Setting $w^S \geq w^D$ implies less labor in S sector: $L_S \leq L_S^D$, more labor in P sector: $L_P \geq L_P^D$, and lower wage for P workers: $w_P \leq w^D < y_w^D$. Because the government can not make transfer to the ruled groups, including P workers, P worker income is always lower than in democracy and P workers do not support oligarchy. Notice this is also related to the setting that the government can only set S sector minimal wage but not the P sector minimal wage. This is realistic for China because the government has better control over state firms and can guarantee that state firms follow the wage regulation and pay high wage but not the private firms.

In the case that w_S is high enough and S workers are supporters of oligarchy, if the number of S workers is sufficiently large, oligarchy is sustained. As we discussed previously, the minimal number of supporters to sustain oligarchy is exogenously give as $\underline{L}$. Later we will discuss what is a reasonable value for $\underline{L}$ in China. If $L_S \leq \bar{L}$, which implies high enough

state wages, and at the same time $L_S \geq \underline{L}$, which guarantees enough supporters, oligarchy is sustained.

To sum up, to sustain oligarchy, the government faces two political constraint. The first is the "*high state wage constraint*", i.e. $w_S \geq y_w^D$ so that S workers support oligarchy. Because equation (1.1) gives a one-to-one mapping from w_S to L_S, we can alternatively treat L_S as the control variable in the high state wage constraint. Then high enough state wage is equivalent to low enough state employment share $L_S \leq \bar{L}$. The second is the "*minimal support constraint*", i.e., $L_S \geq \underline{L}$. So the government faces a critical trade-off between these two political constraints, stated in the following lemma.

Lemma 1.3 (Trade-off of state sector wage)**.** *Increasing w_S guarantees high state wage constraint and buys S workers' political support. However, it reduces S sector employment L_S, which may violate the minimal support constraint.*

The two political constraints give an area of $L_S \in [\underline{L}, \bar{L}]$ that the oligarchy can be sustained. If $\underline{L} \leq \bar{L}$, this area is non-empty, otherwise no L_S can satisfy both constraints. $\underline{L}$ is an exogenous parameter, determined by political power of workers and elites. If citizens are well-organized and have relatively high political power, elites need to buy off many workers to sustain oligarchy. If most citizens are not politically mobilized, the government can stay in power with a small number of supporters. In the latter case, $\underline{L}$ can be low. $\bar{L}$ is endogenously determined by labor allocation in democracy, which again depends on the capital allocation K_S and K_P. $\bar{L}$ is pinned down by $w_S \geq y_w^D$ and can be calculated as follows:

$$\begin{aligned} w_S &= (1-\alpha) K_S^{\alpha} L_S^{-\alpha} \geq y_w^D = \left(1 + \tau^D \frac{\alpha}{1-\alpha}\right)(1-\alpha) K_S^{\alpha} \left(L_S^D\right)^{-\alpha} \Rightarrow \\ L_S &\leq \nu L_S^D = \nu \frac{zK_S}{zK_S + K_P} \doteq \bar{L}, \end{aligned}$$

where $\nu = \left(1 + \tau^D \frac{\alpha}{1-\alpha}\right)^{-\frac{1}{\alpha}}$. So if $\frac{zK_S}{K_P}$ is large enough, $\bar{L}$ can be larger than $\underline{L}$. In other words, sustaining oligarchy requires that S sector is equipped with enough capital, relative to the P sector capital. The equilibrium is summarized in the following proposition.

Proposition 1.4 (Equilibrium given capital allocation)**.** *If there is sufficiently large capital in S sector relative to the capital in P sector, oligarchy can be sustained. In the state sector,*

wage and capital labor ratio are high while capital return is low. P sector capital return and entrepreneur income are higher than in democracy because of low private sector wage. If S sector capital is small, democratization occurs.

The capital labor ratio in S sector is high in oligarchy because of the low S sector labor, as can be seen from 1.4. Because of the high wage, S sector capital return is low. In contrast, because of the low wage and the abundant labor in P sector, P sector capital return is high and entrepreneur income is high.

1.3.3 Discussions on the Equilibrium Given Capital Allocation

Given capital allocation, the government creates a dual labor market: state workers get high wages and hence support the government, while private workers get low wages. This is essentially the so-called "divide-and-rule" strategy: breaking the group of workers into two groups, and providing different economic interests to gain support from one group and maintain power.

The equilibrium given capital allocation are useful to explain three phenomena in current China: (1) large gap of state-private sector wages, (2) middle class's political support for the current regime, (3) higher capital labor ratio and low capital return in the state sector.

The first fact is discussed in section 1.2, and is the immediate consequence of proposition 1.4. High state sector wage is necessary for getting political support from workers, and the general equilibrium effect leads to abundant and cheap labor in the private sector. This contributes to the high inequality among workers. The inequality, provides abundant cheap labor to the private sector, benefits the entrepreneurs. This allows potentially faster capital accumulation and growth of the private sector and the whole economy. We will discuss more on this in the dynamic model.

Second, the middle class, consisting of state workers and entrepreneurs in the model, are supportive to the existing political regime because of the economic benefits. This is consistent with the finding of Chen and Lu (2011) discussed in section 1.2, but on the contrary of the traditional wisdom that the middle class are the natural driving force of democracy, as in the European history. This is not surprising. In the history of democratic movement in Europe, such as the Glorious Revolution and French Revolution, the middle

class were against the aristocracy of the Kings, whose political power relied on repression. The middle class did not rely on the state but emerged from private businesses. In contemporary China, the state sector is large and a significant fraction of the middle class are created by and rely on the state, so they become supporters of the state. It is also similar in many other developing countries. This helps to understand why in some emerging markets, economic growth and the burgeoning bourgeoisie do not push for democratization. For example, The Economist (2009) documents that 95% of adult Kuwaitis work for the government, usually in white-collar civil-service jobs which are typical middle class jobs, while its private-sector middle class consists almost entirely of foreigners. The wage gap between the state and private sector is large there. These distortions keep politically important local workers in the state sector and is a smart way to maintain oligarchy. More examples are the anti-democracy middle class - the "Yellow Shirts" - in Thailand and the growing state middle class linked with growing inequality in 1960's Brazil.

The third fact is well documented in the literature. Song et al. (2011a) show that state sector capital labor ratio is much larger than the private sector, in every industry. Brandt and Zhu (2010) show that the capital return in the state sector is lower than 5% while the number for the private sector is above 50%, as shown in figure 1.5. The difference of capital returns is partly due to the difference of wages and distorted labor allocations. It is also due to the capital allocation, as we will see in the dynamic model below.

In a nutshell, the simple analysis on the equilibrium given capital allocation is useful to illustrate the interactions of the political and economic systems in oligarchy in each period. On the one hand, the political interests, largely shape the state distortions and economic outcome. Taking into account political considerations, we can understand some economic puzzles in China. On the other hand, economic power determines political outcome. When the state sector is economically powerful and equipped with enough capital, elites can keep a large enough supporter base and sustain oligarchy.

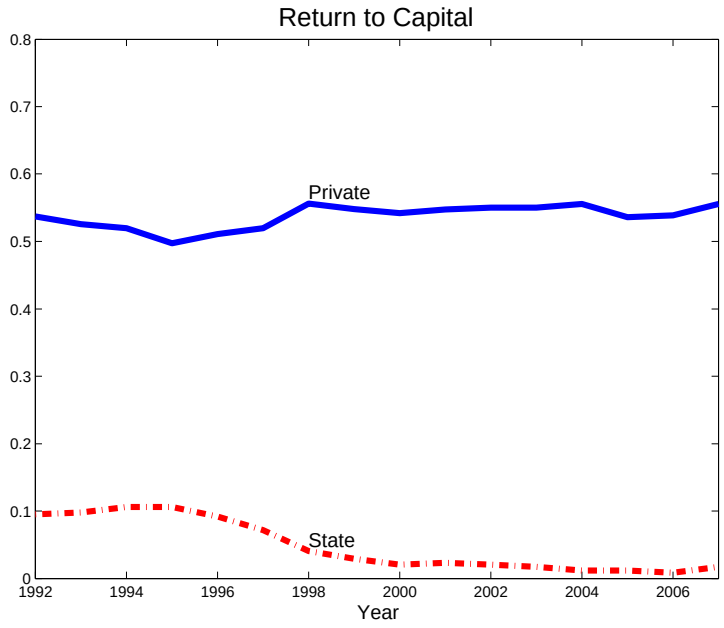

Figure 1.5: Capital return in the state and private sectors.

1.3.4 The Dynamic Equilibrium

In the previous section, we see that the economic power, i.e., size of state capital, relative to the private capital, is crucial for sustaining oligarchy with the "divide-and-rule" strategy. Only when holding abundant state capital, can the government buy enough political support. So the oligarchic government is motivated to control the capital formation and allocation between the state and private sectors. Now we study the whole dynamic equilibrium, including how S and P sector capitals are allocated, tax decisions, consumption and saving in each period, in addition to what we learned in subsection 1.3.2, i.e., the equilibrium given capital allocation.

In democracy, workers control the government to maximize their income. We have seen above that the government does not want to distort the labor market or change the political system. Moreover, it does not impose any financial repression on the more efficient P firms because more capital in P sector implies higher wage for workers. So P firm leverage can reach $\eta = \bar{\eta}$. The government also doesn't want to distort the S sector capital but let it determined by the market if the tax rate is not too high. So we can safely assume that the

equilibrium in democracy is a competitive equilibrium. Finally, the government taxes elites and entrepreneurs to the maximal level $\tau = \bar{\tau}$ and transfers the tax income to workers to maximize worker income in this period. The dynamics in democracy is basically a two sector growth model with an initial misallocation which is removed over time, as in Lewis (1954) and Song et al. (2011a). The dynamic equilibrium is summarized in the following proposition.

Proposition 1.5. *In democracy, elites get return on their asset at interest rate r, and entrepreneur asset yields return greater or equal to r. If β is large enough, entrepreneurs accumulate more and more asset over time. Eventually, the relative size of S sector over P sector, measured by $k = \frac{z_S K_S}{K_P}$, decreases over time to* 0 .

The intuition for the above proposition is the following: efficient labor allocation implies the same wages in S and P sector. S firms compete with each other, so the capital return equals the cost of financing, i.e., the interest rate - r. The capital return pins down S firm capital labor ratio and wage. P firms hire workers at the same wage rate as S firms, but they are more productive, so P firm capital return is higher. S firms exist if P firm capital is small and P firms can't hire all the labor given the wage. This happens when entrepreneur asset is small. In this case, entrepreneurs get higher return than r on their asset. If β is large enough, entrepreneur asset and P firm capital increase over time, and finally P firms hire all workers and S firms all exit. In other words, market force is decisive in this pure economic model without any political constraints.

In oligarchy, the representative elite controls the government and makes four policies - P firm leverage η, S sector capital K_S, minimal wage in S sector w_S, and tax - to maximize her life-time utility. In China, the government controls state banks, and thus is able to determine the size of loans to private firms and affect private firm leverage η. Moreover, the government can use both direct investment and interest subsidy to control state firm capital. In terms of modeling, controlling interest rate for the loan to state firms to influence their capital choice is the same as directly setting the state capital. So in the model we let the government directly set the S sector capital. For the tax, the elites can tax private workers without affecting the political outcome, so the government sets the tax rate on private workers to the maximum. The tax for state workers is set to 0 because taxing them

makes it more difficult to provide them high enough final income and buy their political support. The only tax rate undetermined is the one on private entrepreneurs τ_p.

The decisions of workers and entrepreneurs are simple. S and P workers behave as in subsection 1.3.2, i.e., support oligarchy if and only if the income is higher than in democracy. Then they consume all the income. An entrepreneur, as an infinitely small agent, takes the political outcome and P sector capital return as given. So her choice is simply maximizing the current period income and then consume and save for the next period. The entrepreneur income can be written as:

$$\begin{aligned} y_p &= \max_{K_P} (1+r_P) K_P - R(K_P - a_p), \\ \text{s.t. } K_P &\leq \eta a_p, \end{aligned}$$

where $R = 1+r$, a_p is the entrepreneur asset, and $r_P = (1-\tau_p)\alpha K_P^{\alpha-1} L_P^{1-\alpha} - \delta$ is the P sector capital return. The entrepreneur's choice on K_P is obvious: if $r_p > r$, the entrepreneur chooses to invest as much as possible, the income is proportional to a_p and the return to her asset is larger than r. If P sector capital return is smaller or equal to r, she doesn't invest to the maximal level and gets asset return a_p. Since she lives only on asset return, given the logarithmic utility form, she always saves a constant fraction β of her total income to the next period.

Given our discussion in subsection 1.3.2, we know the equilibrium outcome given K_S, K_P in each period, that is, if $\frac{K_S}{K_P}$ is large enough and w_S is sufficiently high, oligarchy can be sustained. In this case, the one period elite income is the following:

$$y_e = \pi_S + (1-\delta) K_S - R(K_S - a_e) + \bar{\tau} w_P L_P + \tau_p \pi_P, \tag{1.2}$$

where $\pi_S = \alpha (z K_S)^{\alpha} L_S^{1-\alpha}$ and $\pi_P = \alpha (K_P)^{\alpha} L_P^{1-\alpha}$ are capital income of S and P firms.

The representative elite's dynamic problem can be decomposed into two steps. First, she chooses to sustain oligarchy or to democratize:

$$W(a_e, a_p) = \max \left\{ W^O(a_e, a_p), W^D(a_e, a_p) \right\}. \tag{1.3}$$

If the later is chosen, the economy end up in the democratic equilibrium discussed above. If the former is chosen, she picks government policies η, K_S, w_S, τ_p to sustain oligarchy in the second step.[9] Moreover, she decides consumption and saving to maximize her lifetime utility:

$$\begin{aligned} W^O(a_e, a_p) &= \max_{\eta, K_S, w_S, \tau_p, c_e, a'_e} \log c_e + \beta W\left(a'_e, a'_p\right) && (1.4) \\ \text{s.t. } w_S &\geq y_w^D(\eta, K_S, a_p), \\ L_S &\geq \underline{L}, \\ a'_e &= y_e(\eta, K_S, w_S, \tau_p, a_p) - c_e, && (1.5) \\ a'_p &= \beta y_p(\eta, K_S, w_S, \tau_p, a_p). \end{aligned}$$

From the expression of y_e in equation 1.2, we see that within each period a_e only contributes to elites' income through interest revenue, and it does not affect other equilibrium outcomes at all. It also doesn't directly affect future state variables a'_p and a'_e. So the contribution of a_e is simply Ra_e in the elite's budget constraint. Its only role is to smooth the life-time consumption. Therefore the representative elite's problem in oligarchy can be again split into two sub-problems: first, maximization of the lifetime income with discounting rate $\frac{1}{R}$ with government policies; second, maximization of the lifetime utility using a_e to smooth consumption. The second sub-problem is straight-forward and it doesn't affect the first one and other politico-economic outcomes. The first sub-problem has only one state variable, as follows:

$$\begin{aligned} V^O(a_p) &= \max_{\eta, K_S, w_S, \tau_p} \hat{y}_e(K_S, \eta, w_S, \tau_p, a_p) + \frac{1}{R} V\left(a'_p\right) && (1.6) \\ \text{s.t. } w_S &\geq y_w^D(\eta, K_S, a_p), \\ L_S &\geq \underline{L}, \\ a'_p &= \beta y_p(K_S, \eta, w_S, \tau_p, a_p), \end{aligned}$$

[9] Here we assume that if in the first step oligarchy is chosen, then in the second step the government only pick policies that sustain oligarchy. If the government picks policies that can't sustain oligarchy, the economy ends up in democracy in the second step, which gives the same elite income - interest revenue on a_e - as democratizing in the first step. elites are indifferent and assumed to choose to simply democratize in the first step, to save us from infinite equilibria which only differ trivially.

where $\hat{y}_e = y_e - Ra_e$ is the income not related to a_e. Next period value $V\left(a'_p\right)$ depends on the political outcome of next period. We can write:

$$V\left(a_p\right) = \max\left\{V^O\left(a_p\right), V^D\left(a_p\right)\right\},$$

where $V^D\left(a_p\right) = 0$ because the elite income in democracy is simply the asset return Ra_e, as stated in Proposition 1.5.

The model can be solved numerically and the general properties of its solution will be explained in the following parts: (1) given K_S and K_P, the choices of other variables; (2) given K_P, the choice of K_S; (3) the choice of η that affects K_P. Combining the three parts, the structure of the solution will be clear.

First, given K_S and K_P, we know from subsection 1.3.2 that if K_S is large enough, there exists some w_S that sustains oligarchy, or equivalently, some L_S that falls into the region $[\underline{L}, \bar{L}(K_S, K_P)]$. Generally, the optimal choice of w_S is y_w^D, or equivalently, $L_S = \bar{L}$. This choice implies the least labor distortion but still satisfies high state wage constraint.[10] In other words, elites prefer not to distort the labor market more than the necessary. Furthermore, τ_P is generally set at the highest level $\bar{\tau}$.

Then how does the government choose K_S, given K_P? This is the second part of the solution. In figure 1.6, we use a numerical example to depict how state sector labor, political outcome, elite income, and marginal benefit of state capital for elites depend on the choice of K_S (the x-axis). Given a K_P, if K_S reaches certain critical level, there can be enough state workers (left-upper panel) - in this figure, $\underline{L} = 0.2$ - and oligarchy can be sustained (right-upper panel). Then there is a jump of elite income above the critical level of K_S (left-lower panel) because in oligarchy elites control the government and the tax system. For this reason, though the capital return goes down to even lower than 0 as more capital is invested in the state sector, elites still prefer to invest until the critical level of K_S (right-lower panel) to sustain oligarchy.

[10] This is true as long as the tax rate $\bar{\tau}$ is not too high. One sufficient condition is $\bar{\tau} \leq \alpha$, which is a reasonable constraint, considering that α is estimated to be around 0.5 in China. If $\bar{\tau}$ is too large, elites can extract more from the private sector than from the state sector, the solution may change, but this is not very reasonable.

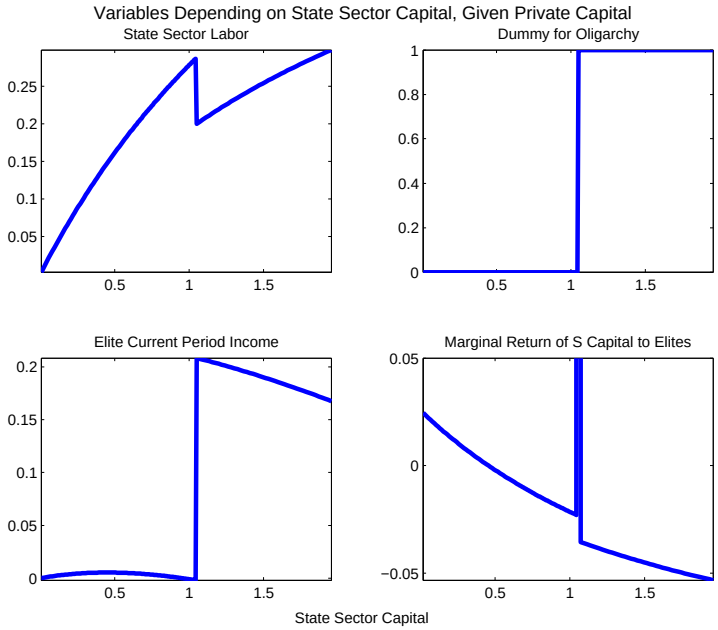

Figure 1.6: The outcome depending on choice of K_S

In the example above, given the particular level of K_P, K_S that just sustains oligarchy gives highest current period income to elites. But for other levels of K_P, the situation may be different. As we can see in figure 1.7, when K_P (the x-axis) is very small, K_S is negatively related to K_P (left-upper panel) and L_S is larger than $\underline{L}$ (right-upper panel).[11] In this region, a larger K_P, corresponds to a larger P sector labor and a smaller S sector labor, hence it is optimal for elites to reduce investment in S sector - K_S - accordingly. However, when K_P is large enough, and S sector labor reaches the minimal level $\underline{L}$, a larger K_P implies that the government has to invest more in S sector to maintain oligarchy. We can see that a larger P sector not only increases benefit for elites - tax income, but also creates higher cost - larger interest payment for K_S (left-lower panel). Due to the decreasing return to capital, there is a level of K_P that maximizes the elite income (right-lower panel).

How do elites set K_P to be closer to the optimal level for them? In the third step here, we discuss the choice of η that affects entrepreneur borrowing ability and capital available for P firms. When the government prefers a larger K_P, it sets $\eta = \bar{\eta}$ and imposes no financial repression. When it wants a smaller K_P, it sets $\eta < \bar{\eta}$, and P firms receive less bank loan

[11] Figure 1.7 comes from the same numerical example as figure 1.6.

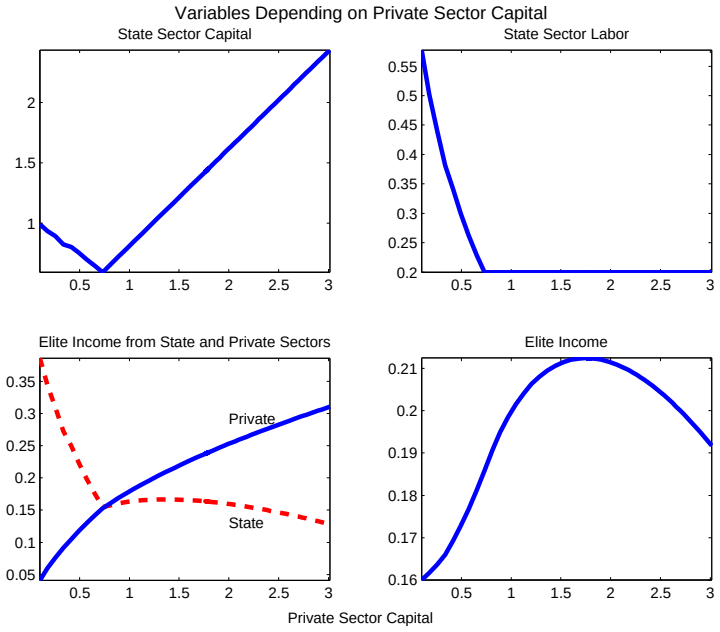

Figure 1.7: The outcome depending on choice of K_S

than the maximal level. This can be seen in figure 1.8.[12] The x-axis is a_p. As we move a_p from very small to very large, the P firm leverage goes down gradually (left-upper panel) as the government prefers K_P not too small or too large. The S sector capital first goes down but then goes up proportionally to the P sector capital (right-upper panel), because enough S employment share needs to be guaranteed (left-lower panel). The government's influence on K_P is limited because η is bounded by $\underline{\eta}$ and $\bar{\eta}$, so it may not be able to set K_P to its favorite level when a_p is too small or too large. This is why the elite lifetime income is maximized for an intermediate level of a_p (right-lower panel). This is the second tradeoff for the elites, in addition to the first tradeoff of state wage and employment.

Lemma 1.6 (Trade-off of private sector capital)**.** *A larger K_p contributes more tax income, but it also requires larger K_S to sustain oligarchy and more interest expense. As K_P increases from a very small level, elites' current-period income first increases and then decreases. elites' lifetime income also follow a similar pattern. This trade-off also applies to entrepreneur asset because it is an important determinant of the private capital.*

[12] Figure 1.8 comes from the same numerical example as figure 1.6.

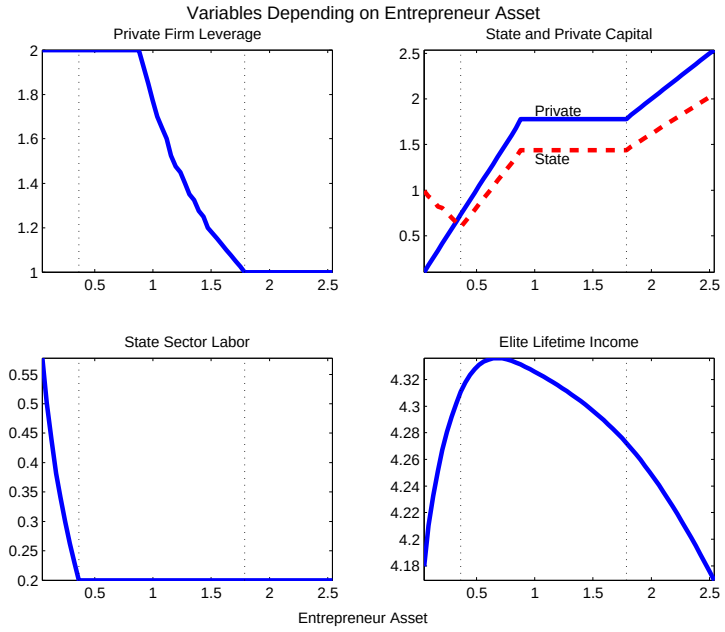

Figure 1.8: Equilibrium variables, depending on entrepreneurs' asset.

Now we have the solution in oligarchy, and the next question is under which conditions does the government choose to democratize or to sustain oligarchy? The government can invest as much as it wants in S sector to guarantee enough state employment with high wage, for any size of P sector capital. However, large investment in S sector means large cost, while the return can be small due to decreasing return to scale. If P sector capital is large enough, sustaining oligarchy gives lower lifetime income to the elite compared to democracy - the line for elites' income in figure (1.8) can drop below the horizontal zero line: $V(a_p) < 0 = V^D(a_p)$ if a_p is large enough. In this case, elites choose to democratize.

Given the solution of equilibrium, we can simulate the dynamics, starting from a small a_p. Will a_p keep growing until $V(a_p) < 0$ and democratization occurs? It depends on the parameter $\underline{L}$. Given other parameters, if $\underline{L}$ is large enough, democratization will occur. In this case, sustaining oligarchy requires many S workers, so elites have to invest a lot in S sector proportional to the P sector capital. As P sector capital grows larger and larger, elites find the cost of maintaining the state sector too large, and it is optimal to democratize for them. However, if $\underline{L}$ is small enough, elites may prefer oligarchy even when the P sector capital reaches its steady state level. Democratization never occurs. So, given small or

large $\underline{L}$, there are two different development paths. The two paths are different in the long-run, but they are similar in the early stages: starting from small P sector, in the beginning, P sector employment share grows until it reaches the critical level for sustaining oligarchy; then the government over-invests in S sector to maintain enough supporters for oligarchy; finally the two paths differ in the long-run. This divergence of two paths is the so-called "critical juncture" of development in Acemoglu and Robinson (2012). The properties of the transition is summarized in the following proposition.

Proposition 1.7 (Three stage transition)**.** *The economy, starting with a small enough private sector, develops along the following path with three stages:*
Stage 1: "Rapid growth". Growth rate is high. Private sector grows rapidly, benefiting from the low wage. Moreover, the government encourages private sector growth and doesn't impose financial repression: $\eta = \bar{\eta}$*. Rapid privatization reallocates labor from the state to the private sector.*
Stage 2: "State capitalism". Growth continues. The government over-invest in the state sector, while restricting private firms' access to the financial market: $\eta < \bar{\eta}$*. Privatization stops and the state employment share stays at the critical level* $\underline{L}$*.*
Stage 3: Two cases.
Case 1: "Middle-income trap". Oligarchy is sustained permanently and growth slows down. State investment keeps growing at the same rate of the private sector capital, to keep state employment share at $\underline{L}$*. Financial repression on private firm reaches the tightest level* $\eta = \underline{\eta}$*. This happens if* $\underline{L}$ *is sufficiently small.*
Case 2: "Sustained growth". Democratization occurs and output growth becomes rapid again. Financial repression and labor market distortion disappear. State sector declines while private sector grows. This happens if $\underline{L}$ *is sufficiently large.*

Figure 1.9 and 1.10 plot key variables and output during the transition in the case of small $\underline{L}$. It ends up at middle-income trap. The three stages are separated by vertical dashed lines. In comparison, we plot the transition in democracy with the blue dashed line, while the transition in oligarchy is the red solid line. Starting in oligarchy, during the first stage, the private sector is small, therefore not a threat to oligarchy. Elites encourage the growth of private capital to extract more tax income. So the government sets $\eta = \bar{\eta}$ to

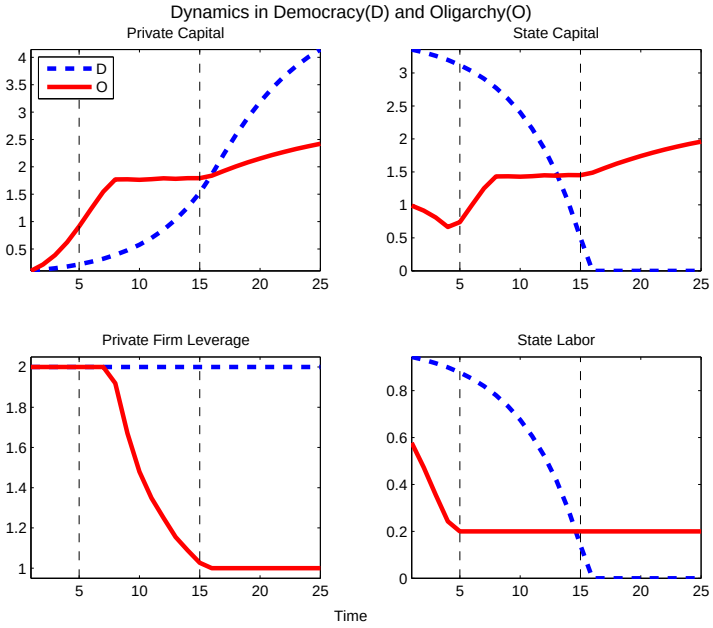

Figure 1.9: Dynamics in democracy (blue) and oligarchy (red) ending in middle-income trap.

lend to private firms as much as possible (left-lower panel of figure 1.9). Moreover, private firms and entrepreneurs benefit from low wage and abundant labor, so private sector capital grows rapidly (left-upper panel). State employment and capital decline accordingly (right-upper panel). Because the more efficient private sector is reallocated with more capital and labor (right-lower panel), the economic growth is rapid (figure 1.10). For this reason, this stage is called "*rapid growth*".

As the private sector grows larger and the state employment share declines to the critical level $\underline{L}$, the economy enters the second stage. The declining state employment share threats the supporter base of oligarchy. If no action is taken, elites can't keep their political power any more. So they increase state investment and then restrict private firms' access to the financial market. Because of the policies in favor of state firms, the state sector keeps its relative economic power and the ability to hire $\underline{L}$ labor with high enough wage. The privatization stops, and no more labor reallocation to the more efficient private sector. However, the large investment in state sector can still keep growth high for a while. But the growth gradually slows down because the financial repression on private firms harms

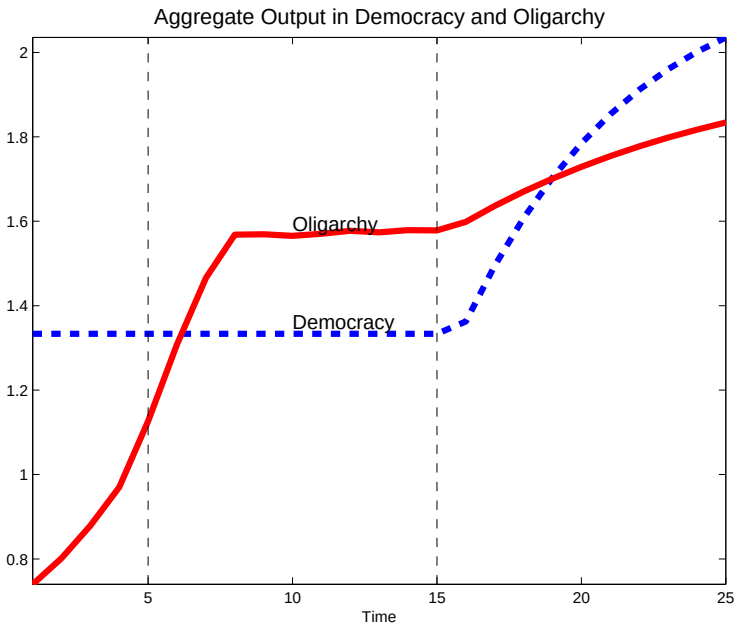

Figure 1.10: Output in democracy (blue) and oligarchy (red) ending in middle-income trap.

the economic efficiency, as shown in the middle section of figure 1.10. This stage features large state investment and financial repression on private firms, so it is a stage of "*state capitalism*". Notice that though the initial output is lower in oligarchy than in democracy, due to the labor market distortion, the output can catch up with democracy in the second stage due to rapid capital accumulation and large state investment.

In the long-run, if $\underline{L}$ is small, elites find it optimal to always sustain oligarchy. They keep over-investing in the state sector as the private sector capital grows to it steady state level. Employment share stays at $\underline{L}$. Though elites have to pay large investment cost, they still extract from tax income from the private sector, so they don't want to democratize. The economy continues as the second stage: over-investment in state firms, financial repression on private firms, no labor reallocation to private firms. The inefficient capital market harms growth. Furthermore, due to decreasing return to capital, growth gradually slows down and eventually output stops growing at the middle level, which is lower than the level in democracy. So in this case, the third stage is called "*middle-income trap*".

If $\underline{L}$ is large enough, elites choose to democratize when the private sector capital reaches certain level. The cost for elites to keep enough workers in the state sector with high

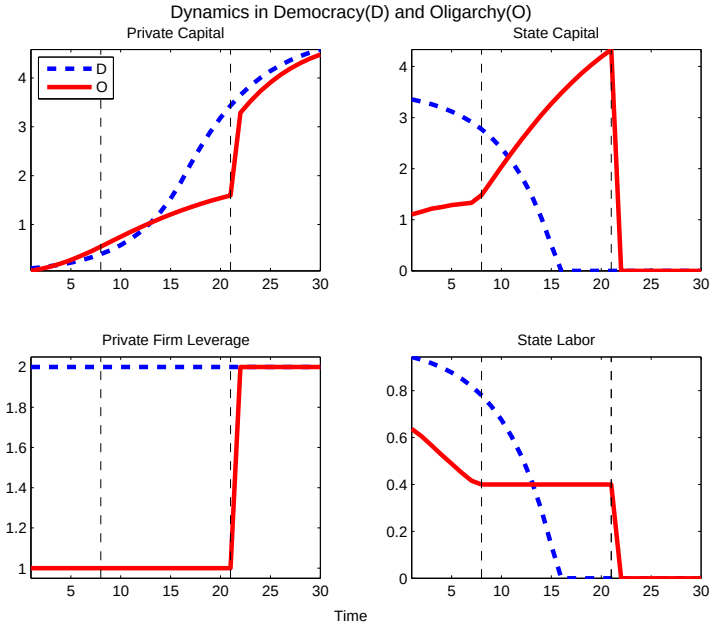

Figure 1.11: Dynamics in democracy (blue) and oligarchy (red) ending in sustained growth.

wage keeps growing as the private sector capital grows. Additionally, marginal return of capital decreases, so elites find the cost of maintaining oligarchy dominates the income in oligarchy when private sector capital grows large enough. They choose to democratize. As we can see in figure 1.11, the state capital quickly drops while the private capital soars up because the financial repression is removed. The output, as shown in figure 1.12, though slightly goes down due to super rapid decline of the state sector, eventually recovers and converges to the high level in democracy.

1.3.5 Quantitative Analysis

Which case of the third stage will be China's future? We calibrate the model to the Chinese economy and provide an answer in this subsection. The targets of the calibration are the key facts in China's recent development, including the wage gap, speed of privatization, and the state employment share.

The economic parameters are set as follows. First, the production function is Cobb-Douglas with the capital share $\alpha = 0.5$ (Bai et al. (2006a)) and depreciation rate $\delta = 0.1$

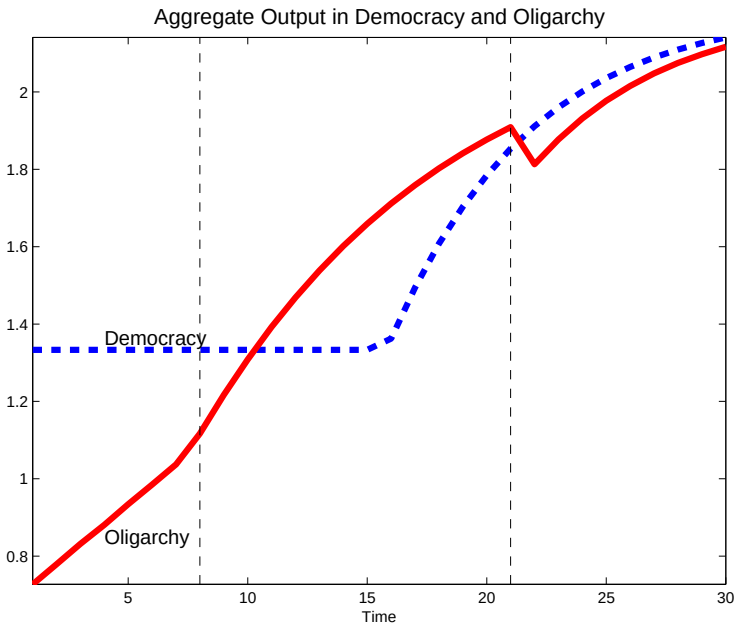

Figure 1.12: Output in democracy (blue) and oligarchy (red) ending in sustained growth.

(Song et al. (2011a)). The state capital efficiency is set to be half of the private capital $z_S = 0.5$. This implies that the TFP of state firms is 71% of the TFP of private firms. This is higher than 59% estimated by Hsieh and Klenow (2009a) with data before 2005, but is reasonable considering that the trend of declining TFP gap discussed in Hsieh and Song (2013). Second, the interest rate of bank saving is $r = 5\%$. Third, the discount factor of entrepreneurs, which is also their saving rate of their lifetime income, is set to $\beta = 0.9$ to match the rapid private sector employment share growth from around 40% to around 80% in 5 years, as we can see from figure 1.2. Finally, the tax rate upper bound is set to $\bar{\tau} = 20\%$ to match the state-private wage gap of 30%, as in figure 1.1. The political parameter in this model is $\underline{L}$, the minimal support needed to sustain oligarchy. We set $\underline{L} = 20\%$, as the state employment share converges to around 20% as in figure 1.2.

Given these parameters, we solve the equilibrium and simulate it starting from a very small private sector: $a_e = 0.05$. Figure 1.8 is the solution of the elites' problem given these parameters, while figure 1.9 is the dynamics (figure 1.11 and 1.12 correspond to setting $\underline{L} = 0.5$). The model's prediction is that China will stay in oligarchy and fall into the middle-income trap, given the current conditions. This is not surprising. The government

is right-now strong, politically and economically, meaning that a relatively small fraction of the citizens' support is sufficient to sustain the current regime, and it has enough financial resource - for example, large foreign reserves - to build up the state sector if it needs to. After 2008 financial crisis, the Chinese government initiates the 400 billion stimulus package and bails out mostly state firms while letting many private firms die. This shows that it keeps the economy and resource allocation under control and stable, and it is able to maintain a powerful state sector to guarantee political stability, according to this theory.

1.3.6 Discussions and Policy Implications

The first two stages of the dynamic model are consistent with China's recent development. From 1997 to around 2003, it is a stage of rapid privatization, as the state employment share declines dramatically. The private sector, in terms of employment share and GDP, grows rapidly, for two reasons. First, the wage is low in the private sector. Compared to state firms which face the regulations on the wage and other payments, including pension tax, health insurance, unemployment insurance and so on, private firms pay relatively low wages, which result in high capital returns. Therefore, private firms accumulate capital rapidly and grow fast. The low wage keeps Chinese private firms competitive. It contributes a lot to the growth of export, and the growth of the economy. Second, the government encourages the private sector growth, because a larger private sector contributes more tax while it is still not too large to threat the supporter base of the government - state employment. So the government encourages various financial resource flowing into the private sector, not only bank loans but also foreign direct investment (FDI), and so on.

At around 2003, as the state employment share approaches the critical level, the privatization slows down and stops dramatically. The direct reason is that more and more investment is diverted to state firms but not private firms. State sector investment share stays at around 60% though its employment is much smaller (see Brandt and Zhu (2010)). The state over-investment retains state employment, but reduces the capital return. In the private sector, the capital return is high, not only because they are more efficient, but also because the credit constraint: private firms can't get enough bank loans for their high return projects. In fact, the financial constraint on private firms has been getting tighter over

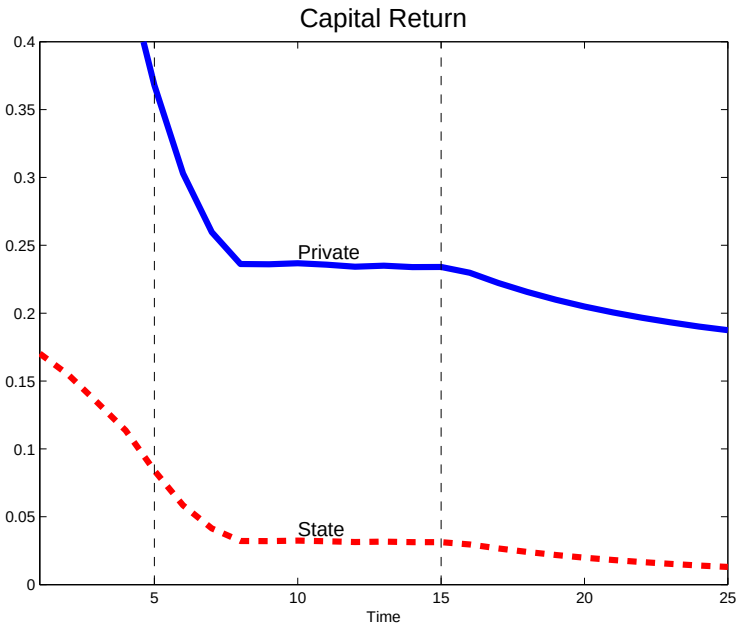

Figure 1.13: Capital return in the state sector (red dashed) and the private sector (blue).

time, signaling growing financial repression on them. The growing financial repression on private firms is formally documented as the growing state-private capital wedge in Brandt and Zhu (2010). The protection on state firms and repression on private firms have gained much attention and are called "the state advances as the private sector retreats". For example, in the passenger airline industry, by 2006, eight private carriers had grew rapidly and had challenged the three state-controlled majors, thanks to the previous government policies encouraging private investors to enter. However, afterwards, the government starts supporting the state airlines and keep them alive with policies including stock purchase from the central government. The state airlines not only survived and also are able to keep their dominance. Our theory's prediction indeed explains why this is happening in the second stage "state capitalism". Elites prefer to maintain a sufficiently strong state sector to guarantee the political control. Our model's prediction on the capital return in the second stage is broadly consistent with the trend: a large gap between the state and private capital returns and declining state capital return, as shown in figure 1.13. Though the capital return in state firms is so slow, the government still keep investing into them to keep them alive.

Is China doomed to fall into the middle-income trap? Not necessary. If the underlining conditions change, the policies and the development path can change accordingly. Mapping into the model, if the parameters such as $\underline{L}, \underline{\eta}, z_S$ change, the government policies and the dynamics, including the third stage, will change. Many policy suggestions on how to switch China's development to a more sustainable path have been made by economists and China watchers. For example, Gary Becker suggested that financial reform should be taken to allocation more resource to private firms, and rural immigrants should be given more rights. Will the government take the suggestions and implement all the policies and reforms that sustain growth? We need to notice that policies or reforms that benefit economic growth may not benefit the elites, who are very influential in the government.

Suppose the government takes a reform that gives more political rights to workers, especially the immigrant workers. This implies that the government has to buy support from a larger fraction of the population. We know that if $\underline{L}$ increases from 0.2 to 0.5 leads to democratization and sustained growth. But do elites like that? Their income goes down to 0 if democratization occurs, so obviously this reform will encounter strong resistance from political elites.

In the above model, we assume that the government is completely under the control of political elites. Some may believe that, in some cases, some technocrats become powerful in the government, and they care only about the output growth in the long-run, but not the economic benefit of elites. In this case, they can initiate reforms which correspond to changing the key parameters of the model, such as $\underline{L}, \underline{\eta}, z_S$. To which extend they can push the reform to depends on their political power in the government, which is modeled as the Nash bargaining power of the following bargaining between technocrats and elites:

$$\max_P \left(Y_\infty(P) - Y_\infty\right)^\alpha \left(V(P) - V\right)^{1-\alpha},$$

where α is the bargaining power of technocrats, $V(P)$ is the lifetime income of elites and $Y_\infty(P)$ is the long-run output given the new parameters after the reform. P can be one of the key parameters $\underline{L}, \underline{\eta}, z_S$. Notice that we consider reform as changing parameters but not the endogenous policy variables such as K_S, η. This implies that technocrats only get a key

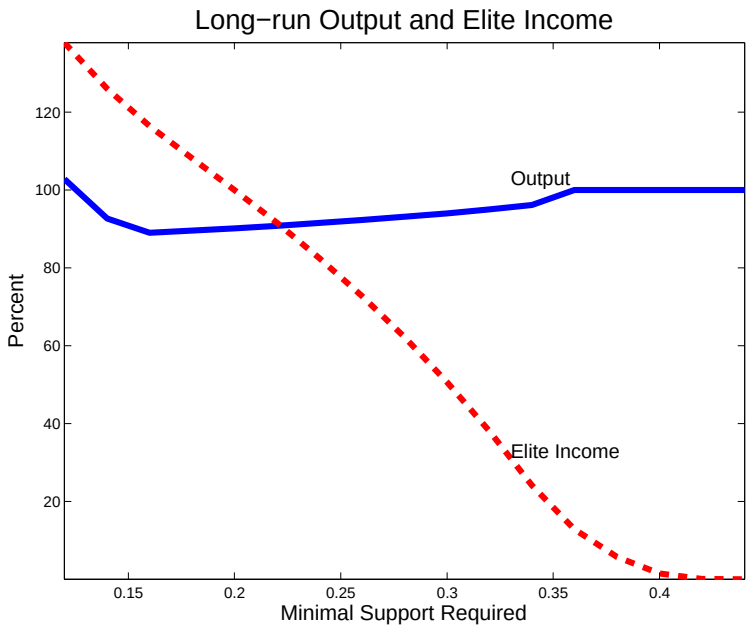

Figure 1.14: Elite income and long-run output depending on support needed.

moment to push for a big change of the society and the political and economic system, and afterwards, the government decisions will be made by elites again.

Figure 1.14 depicts that technocrats would like to increase workers' political rights and increase $\underline{L}$ from the current level $\underline{L} = 0.2$, because this makes the government invest more in the state sector, or even choose to democratize. Both of them lead to larger output levels. However, the reform as the result of the bargaining can only push $\underline{L}$ to the right limited by α. If α is small, the increase of workers' political rights won't be large.

Similarly, financial reform, which reduces the financial repression on private firms can be considered as increasing $\underline{\eta}$. It again increases output, because the private firms can grow larger, and it may even leads to democratization. But again, it harms the elite interests and is hard to be implemented.

One exception is the state firm reform. If a successful reform is taken to increase state firm productivity and reduce the TFP gap between the private and state firms, it increases the output potential. More than that, under the condition that oligarchy is sustained, a more efficient state sector implies that the government can allow the private sector to grow more without worrying about their supporter base - state workers. Less financial repression on

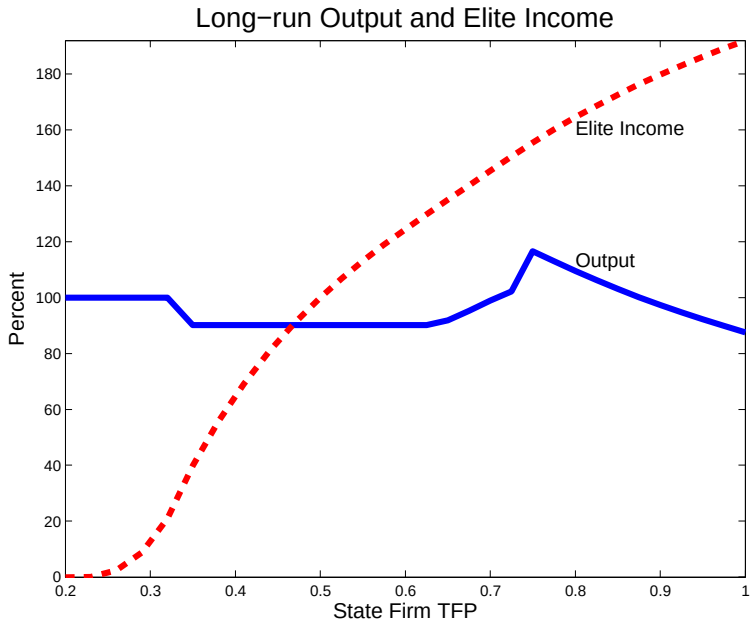

Figure 1.15: Elite income and long-run output depending on state firm TFP.

private firm is needed and higher economic efficiency can be achieved. This reform also increases elite income because of higher total output. Figure 1.15 plots how the long-run efficiency, measured as the long-run output in oligarchy over democracy, can be improved by a more efficient state sector (in the region $z_S \in [0.6, 0.75]$), while the elite income always increase with that. This reform is more likely to be implemented the government. In fact, this is happening right now in China. Hsieh and Song (2013) document the state-private TFP gap is declining. The so-called "industrial upgrading", which aims at building high-tech state firms, is at the top of the agenda for China's further economic reforms. However, it is also very difficult to completely close the gap between the state and private firms, because they are less flexible and provide less economic incentives for the managers, compared to private firms.

1.4 Conclusion

This paper provides a political-economic theory to study China's future economic and political transition and to understand China's recent development. Based on a dynamic growth

model, I add the political constraint that the ruling elite faces: sufficient political supporters. To satisfy the constraint, the government creates a dual labor market, which gives high wages to state workers and turn them into supporters. Moreover, in the financial market, the government encourages private sector growth when it is small enough, but switches to protecting state sector and restricting the private sector when the private sector capital is too large. The economic policies lead to a three-stage transition. The first two stages are "rapid growth" and "state capitalism", which are consistent with a couple of salient aspects of China's development, including (1) rapid growth with repressed wage in the private sector; (2) political support from the middle class, including state sector workers and private entrepreneurs; (3) financial constraints on private firms and support for state firms. In the future, i.e., the third stage of development, China is likely to enter a "middle-income trap" given the current conditions, especially the economically and politically powerful state. To switch to the other development path that leads to "sustained growth", necessary reforms have to be taken, though they may face resistance from elites.

Even though the focus of this paper is on China, it is also useful to study the development of many other emerging countries and even some developed countries with similar patterns compared to China. First, the key political constraint in the theory also exist in some other countries such as Kuwait, Korea in the 80s, and Greece, as political elites or politicians need to buy political support from public workers or workers in industries under their control. So similar stories occur in these countries. Before the 90s, the large local conglomerates (chaebol) in Korea are granted privileged access to low-cost credit. In Kuwait, the oil industry is under the control of the government, so the public sector can hire more than 90% of Kuwaiti nationals with relatively high wage while the private sector is populated with expatriates. Greece public sector workers also receive more than 20% premium (see Giordano et al. (2011)). Second, the theory is also useful to think on a question in development: whether other developing countries should apply the "China model" - the combination of authoritarian politics and state-guided capitalism - to promote economic growth. Some suggestions in favor of adopting this model is based on its past success, but the long-run outcome should be carefully examined and distinguished from the short-run effect. Our theory provides a quantitative framework to evaluate the economic and political consequences.

Further empirical work can be done to examine the theory, especially the three-stage political-economic transition. Which conditions determine the transition to democracy and the long-run growth? Is it consistent with the theory? The theory predicts that if a country can easily build a large state sector - for instance due to rich natural resource - is more likely to sustain the oligarchy, while if efficiency is very important for a country - for example because of exposure of international competition - democratization is more likely to occur. Anecdotal evidence on Gulf countries compared to export oriented economies like Taiwan seem to support the theory. Still, more systematical evidence will be useful to check and improve the theory.

Chapter 2

Sharing High Growth Across Generations: Pensions and Demographic Transition in China[1]

Joint with Zheng Song, Kjetil Storesletten, and Fabrizio Zilibotti

2.1 Introduction

A number of emerging economies are experiencing fast income convergence, improving significantly the average living standards of their populations. Their success is often accompanied by increasing inequality. Intergenerational inequality, which is intrinsically correlated with high economic growth, is an important component of total inequality. Take China, for instance. The present value of earnings for a worker who entered the labor force in 2000 is on average about six times as large as that of a worker who entered in 1970, when China was one of the poorest countries in the world. While young Chinese workers today face much better perspectives than did their parents, poverty among the elderly is pervasive, aggravated by the progressive demise of traditional forms of family insurance (see, e.g., Almås and Johnsen (2013), Park *et al.* 2012, Yang 2011, and Yang and Chen 2010).

[1]This chapter has been submitted to American Economic Journal: Macroeconomics.

In this paper, we study policies and institutional arrangements that enable different generations to share the benefits of high growth in emerging countries. To this aim, we construct a stylized growth model where the economy is initially on a fast convergence trajectory, followed by a slowdown as steady state is approached. A benevolent planner weighs the utility of overlapping generations with a geometrically declining weight. We take as a conservative benchmark a highly forward-looking (*low-discount*) planner who has no desire to redistribute resources across generations in steady state.[2] The planner's goal is to redistribute resources efficiently towards the poorer earlier generations during the transition. We show that the socially desired redistribution can be implemented efficiently by a simple defined benefit pension system with declining replacement rates.

We apply the insights of the theory to the design of the Chinese pension system. To this aim, we extend the stylized model to a multiperiod OLG model based on Song et al. (2011) that can be used for a quantitative analysis. The model embeds key trends of the growth experience of China: a demographic transition, rural-urban migration, fast wage growth – expected to slow down in future, and financial market imperfections driving a gap between the high rate of return to industrial investments and the low rate of return on households' savings. We also introduce into the model a stylized pension system that reproduces salient features of the actual urban Chinese pension system. In line with previous studies, the quantitative analysis shows that the current rules are not financially sustainable, due to the unfavorable demographic transition that will increase sharply the old age dependency ratio.

We consider a number of financially sustainable reforms. The optimal policy entails a significant redistribution from the rich future generations to the current poor generations of Chinese workers. Even a planner with a low annual 0.5% discount rate would like to pay generous pensions to the currently working generations, and negative pensions to subsequent generations. The drive for redistribution would be substantially stronger with a more impatient planner endowed, following Nordhaus (2007), with a social discount rate equal to the market interest rate.

We compare the optimal policy to pension reforms that are being discussed in the policy debate. We start with a sustainable reform involving an immediate permanent reduction in

[2] As we show in section 2.2, a version of this result can be generalized to *any* social discount factor.

the replacement rate, from 60% to 40%, for all workers retiring after 2012.[3] This draconian policy, which we label as the *benchmark reform,* implies the accumulation of a large pension fund until 2051 to pay for the pensions of future generations retiring in times when the dependency ratio will be very high. We show that the benchmark reform entails large welfare losses relative to the optimal policy. While the benchmark reform reduces pensions for the transition generations, the planner would like to increase redistribution towards these groups.

We consider three additional reforms. The first reform is a *delayed reform*, by which the current rules of the Chinese system remain in place until a future date T. Then, benefits are permanently reduced so as to balance the pension system in the long run. The length of the delay is chosen so as to maximize the low-discount planner's utility. The optimal delay is until 2050, so the reform only affects the pensions of workers entering the labor force after 2012. This delay yields large welfare gains for the transition generations relative to the benchmark reform in 2013. Quantitatively, the average gains accruing to the cohorts retiring before 2050 would be equivalent to a 15.3% increase in their lifetime consumption. The generations retiring in 2050 or later would only suffer small losses in the form of a slightly lower replacement ratio.

The second reform is a *fully funded (FF) reform* that replaces the defined benefit transfer-based pension with a fully funded individual account system. To honor existing obligations, the government issues bonds to compensate current workers and retirees for their past contributions. A standard trade-off emerges: all generations retiring after 2059 benefit from the fully funded reform, whereas earlier generations lose. The attractiveness to the planner of this reform is that it reduces tax distortions on labor supply. The shortcoming is that it eliminates a redistributive policy that the planner values. We find that both the low-discount planner and, *a fortiori*, the Nordhaus planner prefer the delayed reform to the FF reform.

The third reform is switching to an unfunded pay-as-you-go (*PAYGO*) system where the replacement rate is endogenously determined by the dependency ratio, subject to a sequence of balanced budget conditions for the pension system. The welfare effects of the

[3]The reform does not renege on the outstanding obligations to current retirees – following the pattern of most reforms in OECD countries.

PAYGO reform are similar, if stronger, to those of a delayed reform. Given the demographic transition of China, the PAYGO yields very generous pensions to early cohorts and punishes severely all generations retiring after 2046. This reform makes the poorer current generations share the benefits of high wage growth with the richer generations that will enter the labor market when China is a mature economy, though it entails larger labor supply distortion than the FF reform.

The results above accrue in an otherwise standard model. We show that in the absence of convergence, e.g., if future wage growth was a constant 2% per year, the model would deliver mainstream predictions: the planner would prefer an FF reform (or, alternatively, the immediate draconian reform) to a delayed reform or to a pure PAYGO system.

The normative predictions of our analysis run against the common wisdom that switching to a pre-funded pension system is the best response for emerging economies facing adverse demographic dynamics. For instance, Feldstein (1999), Feldstein and Liebman (2006) and Dunaway and Arora (2007) argue that a fully funded reform is the best viable option for China. On the contrary, our predictions are aligned with the policy recommendations of Barr and Diamond (2008), arguing against reforming the pension system in the direction of pre-funded individual accounts.[4]

Our results hinge on two common features in emerging economies that are particularly pronounced in China: a high wage growth during transition and a low rate of return on savings (in spite of high returns to investment). In the Chinese case, the forecast of a high wage growth reflects the fact that China's GDP per capita is still below 20% of the US level, leaving ample room for further convergence in technology and productivity. The low rate of return on savings reflect the well-documented fact that China suffers from severe financial market underdevelopment. For instance, Allen *et al.* (2005) document that China has poor investor protection, accounting standards, non-performing loans, etc. relative to its level of development.[5] Our analysis illustrates a point that applies more generally to fast-growing

[4]Barr and Diamond (2008; ch. 15) suggest that (i) although a pre-funded system may induce higher savings (as it does in our model), this objective does not seem valuable for China; (ii) a pre-funded asset-based system is likely to lead to either low pension returns or high risk due to the large imperfections of the Chinese financial system; and (iii) introducing a funded system would benefit future generations of workers at the expense of today's workers who are relatively poor and subject to great economic uncertainty.

[5]Different from us, Feldstein (1999) assumes that the Chinese government has access to a risk-free annual rate of return on the pension fund of 12%. Unsurprisingly, he finds that a fully funded system that collects

emerging economies. Even for economies that are dynamically efficient, the combination of (i) a prolonged period of high wage growth and (ii) a low return on financial savings makes it possible to run a relatively generous pension system over the transition without imposing a large burden to future generations.

The paper is structured as follows. Section 2.2 presents the stylized two-period OLG model from which we derive the normative results that guide the rest of the analysis. Section 2.3 extends it to a partial equilibrium multiperiod OLG model that incorporates the main features of the Chinese pension system. The model is calibrated to the demographic dynamics, the wage and interest rate process (assumed to be exogenous in this section) and the pension system of China. Section 2.4 studies the welfare effects of the alternative pension reforms. Section 2.5 performs sensitivity analysis. Section 2.6 provides a full general equilibrium model of the Chinese economy based on Song et al. (2011), where the wage and interest rate path assumed in section 2.3 are equilibrium outcomes. The general equilibrium model allows us to consider reforms that influence the economic transition. Section 2.7 concludes. The webpage appendix contains some technical material, a description of the Chinese pension system, and additional figures.

2.2 A Model of Intergenerational Redistribution for an Emerging Economy

In this section we lay out a simple dynamic model that illustrates the main point of the paper, namely, that in emerging economies with fast but declining wage growth, even a social planner with a very low discount rate finds it optimal to redistribute resources from future to current generations. Moreover, the optimal redistribution can be implemented by a *simple* pension system that gives higher replacement rates to the earlier generations.

The model economy is populated with two-period lived overlapping generations of households who work when young and live off savings and pension benefits when old. We capture the notion of a temporarily high wage growth (*transition*) by assuming that wages grow at steady rate g from period $t = 1$ and onward, while the growth rate is higher in the

pension contributions and invests these funds at such a remarkable rate of return will dominate a PAYGO pension system that implicitly delivers the same rate of return as aggregate wage growth.

first period: $w_1 > (1+g)\,w_0$. We assume a constant population growth rate n, although this is not essential for the argument. This assumption will be relaxed in the calibrated model of China.

The preferences of a household born in period $t \geq 0$ are given by

$$U_t = \log c_t^Y - \frac{1}{1+\frac{1}{\theta}} h_t^{1+\frac{1}{\theta}} + \beta \log c_{t+1}^O, \tag{2.1}$$

where c^Y and c^O denote consumption in young and old age, respectively, and h denotes labor supply.

Consider a planner endowed with an initial stock of resources (or debt) who can borrow and lend in an international bond market at the gross interest rate R. The planner's resource constraint is given by

$$\sum_{t=0}^{\infty} \left(\frac{1+n}{R}\right)^t \left(c_t^Y + \frac{c_{t+1}^O}{R} - w_t h_t\right) \leq A_0, \tag{2.2}$$

where A_0 denotes the initial planner's wealth net of promises to the initial generation of old. The planner cares about all present and future generations, and discounts the future generations' utilities geometrically, with a discount factor $\phi \in (0,1)$. In order for the resource constraint to be well-defined, we assume that $R > (1+g)\,(1+n)$, i.e., the economy is dynamically efficient. Moreover, we assume that $\phi < (1+n)^{-1}$, so as to ensure that the transversality condition of the planner's problem holds.

The optimal allocation (*first best*) solves the following program:

$$\max_{\left\{c_t^Y, c_{t+1}^O, h_t\right\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \left(\phi\,(1+n)\right)^t \left(\log c_t^Y + \beta \log c_{t+1}^O - \frac{h_t^{1+\frac{1}{\theta}}}{1+\frac{1}{\theta}}\right)$$
$$-\lambda \left(\sum_{t=0}^{\infty} \left(\frac{1+n}{R}\right)^t \left(c_t^Y + \frac{c_{t+1}^O}{R} - w_t h_t\right) - A_0\right).$$

Standard analysis yields:

$$c_t^Y = \lambda^{-1}(\phi R)^t \tag{2.3}$$

$$h_t^{\frac{1}{\theta}} = \frac{w_1}{1+g}\lambda\left(\phi\left(\frac{R}{1+g}\right)\right)^{-t} \quad \text{for } t \geq 1 \tag{2.4}$$

$$h_0^{\frac{1}{\theta}} = \frac{w_1}{1+g}\lambda \times \frac{w_0(1+g)}{w_1}. \tag{2.5}$$

(2.3) is an intergenerational Euler equation reflecting the planner's preferences for redistribution. (2.4)–(2.5) are optimal labor supply conditions. Note that λ, the Lagrange multiplier of the government budget constraint, is an inverse function of the planner's initial wealth, A_0.

Next, suppose that the planner faces a standard implementability constraint: any (Ramsey) allocation must be a competitive equilibrium. Suppose, in addition, that the only instrument at her disposal is a pension system comprising a sequence of taxes and pension replacement rates $\{\zeta_t, \tau_t\}_{t=0}^{\infty}$, where labor income of generation t is taxed at the flat rate τ_t, and the generation receives a pension p_t proportional to aggregate after-tax labor earnings:[6]

$$p_{t+1} = \zeta_t(1-\tau_t)w_t\bar{h}_t,$$

where $\bar{h}_t$ is aggregate labor supply. The resulting lifetime budget constraint for the agents is:

$$h_t(1-\tau_t)w_t + \frac{p_{t+1}}{R} = c_t^Y + \frac{c_{t+1}^O}{R}. \tag{2.6}$$

Solving the household problem, (2.1), subject to (2.6), and recalling that in equilibrium $h_t = \bar{h}_t$, yields the equilibrium labor supply

$$h_t = (1+\beta)^{\frac{\theta}{1+\theta}}\left(1+\frac{\zeta_t}{R}\right)^{-\frac{\theta}{1+\theta}}. \tag{2.7}$$

Moreover, the lifetime consumption growth satisfies a standard Euler equation: $c_{t+1}^O = \beta R c_t^Y$.

[6]The restriction that pension payments are related to average rather than individual labor supply is without loss of generality since, as we show, this policy implements the first best.

The following proposition (proof in the Appendix) establishes that the first best can be implemented by setting tax to zero and choosing a suitable replacement rate sequence.[7]

Proposition 2.1. *The first best allocation is implemented by setting* $\tau_t = 0$ *and the replacement rate sequence*

$$\zeta_t = R\left(\left(\left(\phi\frac{R}{1+g}\right)^t \frac{(1+\theta)(1+g)}{\lambda w_1}\right)^{1+\theta} - 1\right) \qquad \text{for all } t \geq 1, \tag{2.8}$$

and

$$\zeta_0 = R\left(\left(\frac{(1+\theta)(1+g)}{\lambda w_1}\right)^{1+\theta}\left(\frac{w_1}{w_0(1+g)}\right)^{1+\theta} - 1\right), \tag{2.9}$$

where λ *(which is strictly decreasing in* A_0*) is the Lagrangian multiplier associated with the planner's budget constraint.*

Note that, for $t \geq 1$, the replacement rate sequence may increase or decrease over time depending on whether $\phi \gtreqless (1+g)/R$. In particular, if $\phi = (1+g)/R$, i.e., the discount rate is equal to the growth-adjusted net interest rate, then, the planner has no drive for intergenerational redistribution in steady state.[8] Under this particular discount factor, which is an interesting benchmark for the quantitative analysis below, the optimal policy is particularly simple.

Corollary 2.2. *If the planner's discount factor is* $\phi = (1+g)/R$*, then the pension replacement rate is constant for* $t \geq 1$ *(i.e.,* $\zeta_t = \zeta$*). The replacement rate of the transition generation is larger than that of the future generations (*$\zeta_0 > \zeta$*). As a result, for* $t \geq 1$*,* $h_t = \bar{h} \equiv \left(\frac{w_1}{1+g}\lambda\right)^{\theta}$ *and* $h_0 < \bar{h}$ *(see equations (2.4)-(2.5)).*

Intuitively, as the transition generation is especially poor, even a utilitarian planner with no desire to redistribute in steady state would bail out the first generation with a generous pension.[9] In addition, while in the laissez-faire equilibrium all generations supply the

[7]The proof is straightforward. It amounts to verify that the pension policy, (2.8)-(2.9), implies individual consumption and labor supply consistent with the first best allocation (2.3)-(2.5).

[8]Note also that this planner's discount rate is consistent with the standard calibration of infinite-horizon growth models. In particular, the standard steady-state condition of a discrete-time models yields exactly a discount factor of $(1+g)/R$.

[9]More generally, conditional on any planner's discount factor that may imply a trend in pension payments, the first best is implemented by paying a higher pension to the transition generation.

same labor effort, the planner wants the less productive first generation to work less. To achieve these goals, the planner has two distortionary instruments at her disposal: τ and ζ. Interestingly, she finds it optimal not to use labor taxes ($\tau = 0$). The reason is that τ does not affect labor supply (due to a cancellation of an income and a substitution effect), whereas an appropriately designed pension sequence attains both the desired resource redistribution and the labor supply decision that the planner wishes to implement.[10]

In Proposition 2.1, the steady state replacement rate may be negative if the initial wealth, A_0, is low (i.e., if λ is high, see equations (2.8)-(2.9)). Since it may be difficult, in practice, to pay negative pensions, it is interesting to study the case in which replacement rates are constrained to be non-negative. Consider, again, the planning weight $\phi = (1+g)/R$. Also, we focus on an intermediate range of λ – and, thus, of A_0 – such that, on the one hand, the non-negativity constraint on pensions is binding, i.e., the planner is not sufficiently rich to achieve the first best by setting zero taxes in all periods, and paying strictly positive benefits to all generations. On the other hand, the planner is sufficiently rich not to be forced to tax all agents, including the first generation, and pay no pensions at any time (a sufficient condition is that $A_0 \geq 0$).[11] Proposition 2.3 establishes (proof in the Appendix) that the Ramsey planner sets $\tau_0 = 0$ and pay a positive pension to the transition generation, and sets positive taxes and pay zero pensions to all subsequent generations.

Proposition 2.3. *Suppose that $\phi = (1+g)/R$ and that the planner cannot set negative replacement rates ($\zeta_t \geq 0$, for all $t \geq 0$). Assume further, that $\lambda \in [\lambda_1, \lambda_2]$, where $\lambda_1 \equiv (1+\beta)^{\frac{1}{1+\sigma}} \frac{1+g}{w_1}$ and $\lambda_2 \equiv \frac{1}{w_0}(1+\beta)^{\frac{1}{1+\sigma}}$. Then, the Ramsey constrained (second best) allocation is implemented by setting the following tax sequence:*

$$\tau_0 = 0, \tag{2.10}$$

$$\tau_t = 1 - \frac{1+g}{\lambda w_1}(1+\beta)^{\frac{1}{1+\theta}} \geq 0 \qquad \text{for all } t \geq 1, \tag{2.11}$$

[10] This result extends to any preferences featuring a wealth effect to labor supply.

[11] Extending the Proposition to cases of large and small A_0 is straightforward, We focus on the interesting case for simplicity.

and the following replacement rate sequence:

$$\zeta_0 = R\left(\left(\lambda w_0\right)^{-(1+\theta)}\left(1+\beta\right)-1\right), \tag{2.12}$$

$$\zeta_t = 0 \qquad \textit{for all } t \geq 1, \tag{2.13}$$

where λ is the Lagrangian multiplier associated with the planner's budget constraint (where λ is strictly decreasing in A_0).

Intuitively, the planner should never tax and pay benefits to the same generation, as this creates an unnecessary labor supply wedge that can be eliminated by cutting taxes and setting benefits to zero for that cohort. It is easy to show that a version of this result extends to general ϕ.

The results established so far extend to a model in which agents live for $J > 2$ periods and retire after $J_W \geq 1$ periods, and where the high-growth transition lasts for $\tilde{T} \geq 1$ periods (see Propositions B.1 and B.3 and Corollary B.2in the Appendix). Here, we summarize the main results. First, the planner's optimal choice prescribes that the consumption and labor supply sequences of an individual born in period t of age j be as follows:

$$c_{t,0} = \lambda^{-1}\left(\phi R\right)^t, \quad c_{t,j} = c_{t,0}\left(\beta R\right)^j, \quad h_{t,j\leq J_w} = \left(\frac{w_{t+j}}{c_{t,j}}\right)^{\theta}.$$

As in the two-period model, the planner with a discount factor $\phi = (1+g)/R$ wants consumption to grow across cohorts at the rate g implying no redistribution in the steady state, and yet desires positive redistribution in favor of the earlier cohorts which live through the high-growth transition. Second, the first-best allocation can be implemented by a sequence of cohort-specific pension benefits, whereas taxes are set to zero in all periods. Pensions replacement rates decline during the transition, and become constant (possibly, negative) in the steady state. Third, if negative pensions are ruled out, the second best allocation implies that some cohorts pay no taxes and receive pensions, whereas other cohorts pay taxes and receive no pensions. The replacement rate declines across cohorts that live part of their lives during the transition, and part in the steady state.[12]

[12]Note that the optimal tax sequence for a cohort is constant over the life cycle. This is a common feature of models with homothetic preferences, as shown by Garriga (1999).

2.3 A Quantitative Multiperiod Model for China

Guided by the insights of the simple model laid out in section 2.2, we now move to the main contribution of the paper, namely, to construct a quantitative model aimed to evaluate the welfare implications of alternative pension reforms of China. In this section, we lay out a multiperiod OLG model close in spirit to Auerbach and Kotlikoff (1987), Conesa and Garriga (2008), Conesa and Krueger (1999), Huang, Imrohoglu and Sargent (1997), and Storesletten (2000). Then, we specify demographic dynamics, an exogenous wage growth process and a set of pension rules. Finally, we use the model to assess the welfare effects of alternative pension reforms.

2.3.1 Household

The model economy is populated by a sequence of overlapping generations of agents. Each agent lives up to $J - J_C$ years and has an unconditional probability of surviving until age j equal to s_j. During their first $J_C - 1$ years (childhood), agents are economically inactive, make no choices, and gain no utility. Preferences are defined over consumption and leisure and are represented by a standard lifetime utility function,

$$U_t = \sum_{j=0}^{J} s_j \beta^j u\left(c_{t+j}, h_{t+j}\right),$$

where β is the discount factor, c is consumption, and h is labor supply. Here, t denotes the period in which the agent becomes adult (i.e., economically active). Thus, U_t is the discounted utility of an agent born in period $t - J_C$.

Workers are active until age J_W. For simplicity, we abstract from an endogenous choice of retirement. Incorporating endogenous retirement would require a more sophisticated model of labor supply, including non-convexities in labor market participation and declining health and productivity in old age (see, e.g., Rogerson and Wallenius, 2009). Since China has a mandatory retirement policy, the assumption of exogenous retirement seems reasonable. After retirement, agents receive pension benefits until death. Wages are subject to proportional taxes. Adult workers and retirees can borrow and deposit their savings

with banks paying a gross annual interest rate R. A perfect annuity market allows agents to insure against uncertainty about the time of death.

Agents maximize U_t, subject to a lifetime budget constraint,

$$\sum_{j=0}^{J} \frac{s_j}{R^j} c_{t+j} = \sum_{j=0}^{J_W} \frac{s_j}{R^j} \left(1 - \tau_{t+j}\right) \zeta_j \eta_t w_{t+j} h_{t,t+j} + \sum_{j=J_W+1}^{J} \frac{s_j}{R^j} b_{t,t+j},$$

where $b_{t,t+j}$ denotes the pension benefit accruing in period $t+j$ to a person who became adult in period t, w_{t+j} is the wage rate per efficiency unit at $t+j$, η_t denotes the human capital specific to the cohort turning adult in t (we abstract from within-cohort differences in human capital across workers), τ_t is the labor income tax in period t, and ζ_j is the efficiency units per hour worked for a worker with j years of experience, which captures the experience-wage profile.

2.3.2 Demographic Model

Since China faces a major demographic transition that affects the financial sustainability of the pension system, we construct in this section a detailed demographic model with exogenous population dynamics emphasizing internal migration patterns.

Throughout the 1950s and 1960s, the total fertility rate (henceforth, TFR) of China was between five and six. High fertility, together with declining mortality, brought about a rapid expansion of the total population. The 1982 census estimated a population size of one billion, 70% higher than in the 1953 census. The view that a booming population is a burden on the development process led the government to introduce measures to curb fertility during the 1970s, culminating in the one-child policy of 1978. This policy imposes severe sanctions on couples having more than one child. The policy underwent a few reforms and is currently more lenient to rural families and ethnic minorities. Today's TFR is below replacement level, although there is no consensus about its exact level. Estimates based on the 2000 census and earlier surveys range between 1.5 and 1.8 (e.g., Zhang and Zhao, 2006). Recent estimates suggest a TFR of about 1.6 (see Zeng, 2007).

2.3.2.1 Natural Population Projections

We consider, first, a model without rural-urban migration, which is referred to as the *natural* population dynamics. We break down the population by birth place (rural vs. urban), age, and gender. The initial population size and distribution are matched to the adjusted 2000 census data.[13] There is consensus among demographers that birth rates have been underreported, causing a deficit of 30 to 37 million children in the 2000 census.[14] To heed this concern, we take the rural-urban population and age-gender distribution from the 2000 census – with the subsequent National Bureau of Statistics (NBS) revisions – and then amend this by adding the missing children for each age group, according to the estimates of Goodkind (2004).

The initial group-specific mortality rates are also estimated from the 2000 census, yielding a life expectancy at birth of 71.1 years, which is very close to the World Development Indicator figure in the same year (71.2). Life expectancy is likely to continue to increase as China becomes richer. Therefore, we set the mortality rates in 2020, 2050, and 2080 to match the demographic projection by Zeng (2007) and use linear interpolation over the intermediate periods. We assume no further change after 2080. This implies a long-run life expectancy of 81.9 years.

The age-specific urban and rural fertility rates for 2000 and 2005 are estimated using the 2000 census and the 2005 one-percent population survey, respectively. We interpolate linearly the years 2001-2004, and assume age-specific fertility rates to remain constant at the 2005 level over the period 2006-2012. This yields average urban and rural TFR's of 1.2 and 1.98, respectively.[15] Between 2013 and 2050, we assume age-specific fertility rates to remain constant in rural areas. This is motivated by the observation that, according to

[13]The 2000 census data are broadly regarded as a reliable source (see, e.g., Lavely, 2001; Goodkind, 2004). The total population was originally estimated to be 1.24 billion, later revised by the NBS to 1.27 billion (see the Main Data Bulletin of 2000 National Population Census). The NBS also adjusted the urban-to-rural population ratio from 36.9% to 36%.

[14]See Goodkind (2004). A similar estimate is obtained by Zhang and Cui (2003), who use primary school enrolments to back out the actual child population.

[15]The acute gender imbalance is taken into account in our model. However, demographers view it as unlikely that such imbalance will persist at the current high levels. Following Zeng (2007), we assume that the urban gender ratio will decline linearly from 1.145 to 1.05 from 2000 to 2030, and that the rural gender imbalance falls from 1.19 to 1.06 over the same time interval. No change is assumed thereafter. Our results are robust to plausible changes in the gender imbalance.

the current legislation, a growing share of urban couples (in particular, those in which each spouse is an only child) will be allowed to have two children. In addition, some provinces are discussing a relaxation of the current rule, that would allow even urban couples in which only one spouse is an only child to have two children. Zeng (2007) estimates that such a policy would increase the urban TFR from 1.2 to 1.8 (*second scenario* in Zeng, 2007). Accordingly, we assume that the TFR increases to 1.8 in 2013 and then remains constant until 2050.

A long-run TFR of 1.8 implies an ever-shrinking population. We follow the United Nations population forecasts and assume that in the long run the population will be stable. This requires that the TFR converges to 2.08, which is the reproduction rate in our model, in the long run. In order to smooth the demographic change, we assume that both rural and urban fertility rates start growing in 2051, and we use a linear interpolation of the TFRs for the years 2051-2099. Since long-run forecasts are subject to large uncertainty, we also consider an alternative scenario with lower fertility.

2.3.2.2 Rural-urban Migration

Rural-urban migration has been a prominent feature of the Chinese economy since the 1990s. There are two categories of rural-urban migrants. The first category is all individuals who physically move from rural to urban areas. It includes both people who change their registered permanent residence (i.e., *hukou* workers) and people who reside and work in urban areas but retain an official residence in a rural area (*non-hukou* urban workers).[16] The second category is all individuals who do not move but whose place of registered residence switches from being classified as rural into being classified as urban.[17] We define the sum of the two categories as the net migration flow (NMF).

[16]There are important differences across these two subcategories. Most non resident workers are currently not covered by any form of urban social insurance including pensions. However, some relaxation of the system has occurred in recent years. The system underwent some reforms in 2005, and in 2006 the central government abolished the hukou requirement for civil servants (Chan and Buckingham, 2008). Since there are no reliable estimates of the number of non-hukou workers, and in addition there is uncertainty about how the legislation will evolve in future years, we decided not to distinguish explicitly between the two categories of migrants in the model. This assumption is of importance with regard to the coverage of different types of workers in the Chinese pension system. We return to this discussion below.

[17]This was a sizeable group in the 1990s: according to *China Civil Affairs Statistical Yearbooks*, a total of 8,439 new towns were established from 1990 to 2000 and 44 million rural citizens became urban citizens

We propose a simple model of migration where the age- and gender-specific emigration rates are fixed over time. Although emigration rates are likely to respond to the urban-rural wage gap, pension and health care entitlements for migrants, the rural old-age dependency ratio, and so on, we will abstract from this and maintain that the demographic development only depends on the age distribution of rural workers. It is generally difficult, even for developed countries, to predict the internal migration patterns (see, e.g., Kaplan and Schulhofer-Wohl, 2012). In China, pervasive legal and administrative regulations compound this problem.

We start by estimating the NMF and its associated distribution across age and gender. This estimation is the backbone of our projection of migration and the implied rural and urban population dynamics. We use the 2000 census to construct a projection of the *natural* rural and urban population until 2005 based on the method described in section 2.3.2.1. We can then estimate the NMF and its distribution across age groups by taking the difference between the 2005 projection of the *natural* population and the realized population distribution according to the 2005 survey.[18] The technical details of the estimation can be found in Appendix B.2.

According to our estimates, the overall NMF between 2001 and 2005 was 88 million, corresponding to 10.8% of the rural population in 2000.[19] Survey data show that the urban population grows at an annual 4.1% rate between 2000 and 2005. Hence, 89% of the Chinese urban population growth during those years appears to be accounted for by rural-urban migration. Our estimate implies an annual flow of 17.6 million migrants between

(Hu, 2003). However, the importance of reclassified areas has declined after 2000. Only 24 prefectures were reclassified as prefecture-level cities in 2000-2009, while 88 prefectures were reclassified in 1991-2000.

[18]Our method is related to Johnson (2003), who also exploits *natural* population growth rates. Our work is different from Johnson's in three respects. First, his focus is on migration across provinces, whereas we estimate rural-urban migration. Second, Johnson only estimates the total migration flow, whereas we obtain a full age-gender structure of migration. Finally, our estimation takes care of measurement error in the census and survey (see discussion above), which were not considered in previous studies.

[19]There are a number of inconsistencies across censuses and surveys. Notable examples include changes in the definition of city population and urban area (see, e.g., Zhou and Ma, 2003; Duan and Sun, 2006). Such inconsistencies could potentially bias our estimates. In particular, the definition of urban population in the 2005 survey is inconsistent with that in the 2000 census. In the 2000 census, urban population refers to the resident population (*changzhu renkou*) of the place of enumeration who had resided there for at least six months on census day. The minimum requirement was removed in the 2005 survey. Therefore, relative to the 2005 survey definition, rural population tends to be over-counted in the 2000 census. This tends to bias our NMF estimates downward.

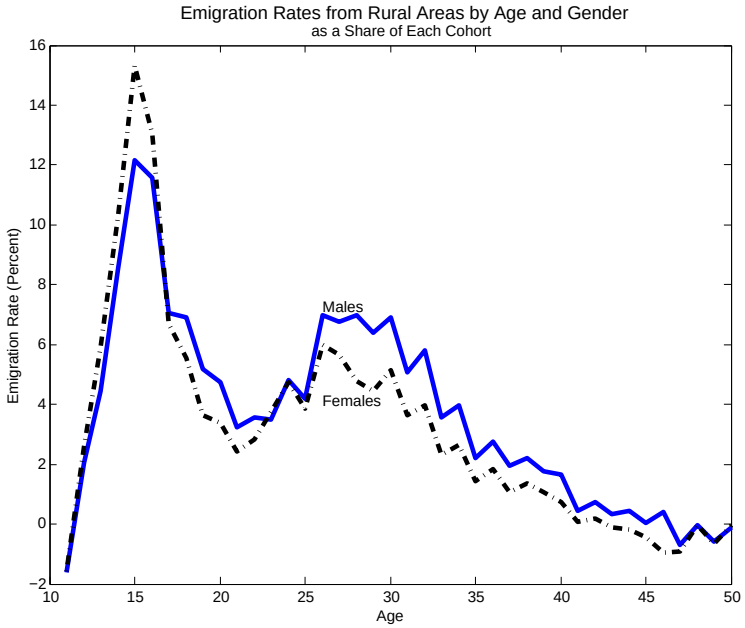

Figure 2.1: Emigration rates

The figure shows rural-urban migration rates by age and gender as a share of each cohort. The estimates are smoothed by five-year moving averages.

2001 to 2005, equal to an annual 2.3% of the rural population. This figure is in line with estimates of earlier studies. For instance, Hu (2003) estimates an annual flow between 17.5 and 19.5 million in the period 1996–2000.

The estimated age-gender-specific migration rates are shown in figure 2.1. Both the female and male migration rates peak at age fifteen, with 15.3% for females and 12.2% for males. The migration rate falls gradually at later ages, remaining above 1% until age thirty-nine for females and until age forty for males. Migration becomes negligible after age forty.

To incorporate rural-urban migration in our population projection, we make two assumptions. First, the age-gender-specific migration rates remain constant after 2005 at the level of our estimates for the period 2000–2005. Second, once the migrants have moved to an urban area, their fertility and mortality rates are assumed to be the same as those of urban residents.

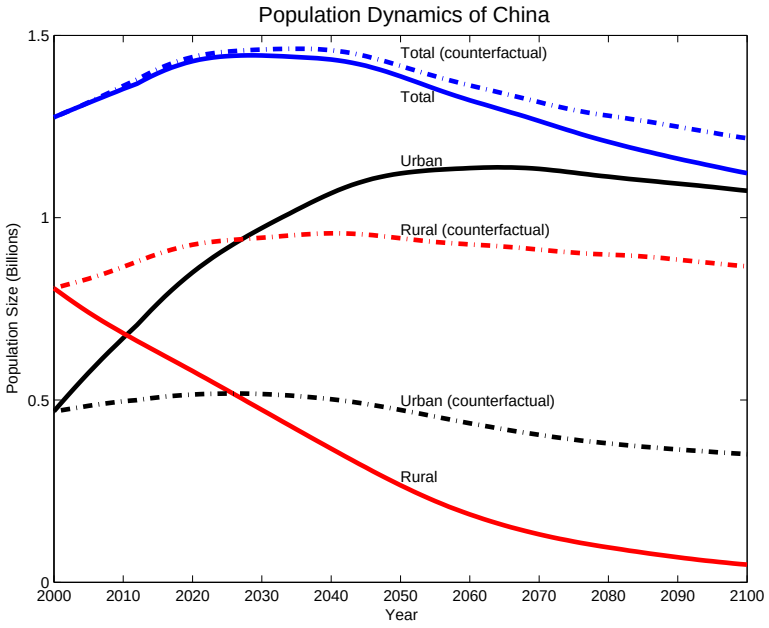

Figure 2.2: Population dynamics

The figure shows the projected population dynamics for 2000-2100 (solid lines) broken down by rural and urban population. The dashed lines show the corresponding natural population dynamics (i.e., the counterfactual projection under a zero urban-rural migration scenario.

Figure 2.2 shows the resulting projected population dynamics (solid lines). For comparison, we also plot the natural population dynamics (i.e., the population model without migration [dotted lines]). The rural population declines throughout the whole period. The urban population share increases from 51% in 2011 to 81% in 2050 and to over 95% in 2100. In absolute terms, the urban population increases from 450 million in 2000 to its long-run 1.2 billion level in 2050. Between 2050 and 2100 there are two opposite forces that tend to stabilize the urban population: on the one hand, fertility is below replacement in urban areas until 2100; on the other hand, there is still sizeable immigration from rural areas. In contrast, had there been no migration, the urban population would have already started declining in 2008.

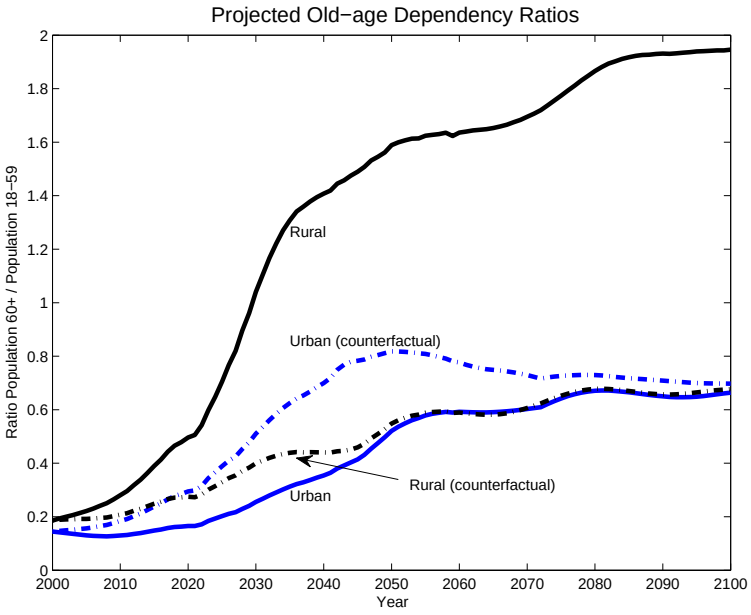

Figure 2.3: Dependency ratio

The figure shows the projected old-age dependency ratios, defined as the ratio of population 60+ over population 18-59, for 2000-2100 (solid lines) broken down on urban and rural population. The dashed lines show the corresponding ratios under the zero migration counterfactual (i.e., the natural population dynamics).

Figure 2.3 plots the old-age dependency ratio (i.e., the number of retirees as percentage of individuals in working age [18-60]) broken down by rural and urban areas (solid lines).[20] We also plot, for contrast, the old-age dependency ratio in the no migration counterfactual (dashed lines). Rural-urban migration is very important for the projection. The projected urban old-age dependency ratio is 52% in 2050, but it would be as high as 82% in the no migration counterfactual. This is an important statistic, since the Chinese pension system only covers urban workers, so its sustainability hinges on the urban old-age dependency ratio.

[20] In China, the official retirement age is 55 for females and 60 for males. In the rest of the paper, we ignore this distinction and assume that all individuals retire at age 60, anticipating that the age of retirement is likely to increase in the near future. We also consider the effect of changes in the retirement age.

2.3.3 Calibration

In this section, we calibrate the wage process and other key parameters of the model to China. One period is defined as a year and agents can live up to 100 years ($J = 100$). The demographic process (mortality, migration, and fertility) is described in section 2.3.2. Agents become adult (i.e., economically active) at age $J_C = 22$ and retire at age 60, which is the male retirement age in China (so $J_W = 59$). Hence, workers retire after 38 years of work.

The wage growth process is taken as exogenous. In section 2.6 we show that the assumed wage process is the equilibrium outcome of a calibrated dynamic general-equilibrium model with credit market imperfections close in spirit to Song et al. (2011). We set the age-wage profile $\left\{\zeta_j\right\}_{j=23}^{59}$ equal to the one estimated by Song and Yang (2010) for Chinese urban workers. This implies an average return to experience of 0.5%. In this section of the paper, we take the hourly wage rate as exogenous. The assumed dynamics of urban wages per effective unit of labor is shown in figure 2.4: Hourly wages (conditional on human capital) grow at approximately 5.7% between 2000 and 2011. This is in line with the estimate of Ge and Yang (2013) who document that the wage of workers with only middle school education grew by 5.9% over the 1992-2007 period. For the future, we assume an annual growth of 5.1% between 2011 and 2030, subsequently declining to 2.7% between 2030 and 2050. In the long run, wages are assumed to grow at 2% per year, in line with wage growth in the United States over the last century. In section 2.6, we show that the assumed wage rate dynamics of figure 2.4 is the equilibrium outcome of a calibrated version of the model of Song et al. (2011).

There has been substantial human capital accumulation in China over the last two decades. To incorporate this aspect, we assume that each generation has a cohort-specific education level, which is matched to the average years of education by cohort according to Barro and Lee (2010). The values for cohorts born after 1990 are extrapolated linearly, assuming that the growth in the years of schooling ceases in year 2000 when it reaches an average of 12 years, which is the current level for the US. We assume an annual return of 10% per year of education.[21] Since younger cohorts have more years of education, wage

[21] Zhang et al. (2005) estimated returns to education in urban areas of six provinces from 1988 to 2001. The average returns were 10.3% in 2001.

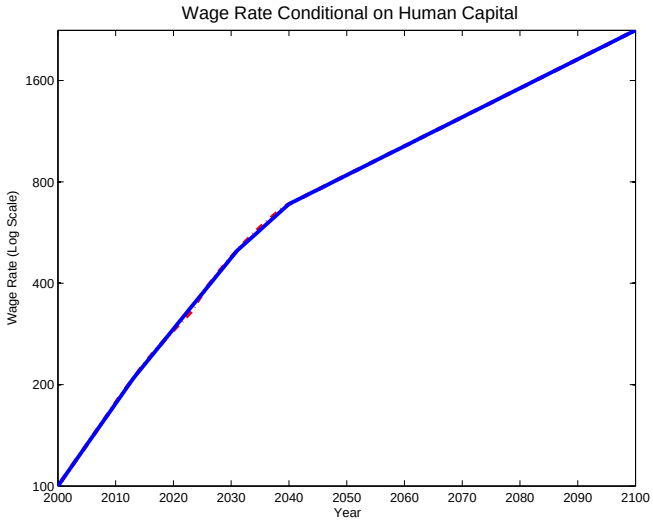

Figure 2.4: Wage

The figure shows the projected hourly wage rate per unit of human capital in urban areas, normalized to 100 in 2000. The process is the endogenous outcome of the general equilibrium model of section 2.6.

growth across cohorts will exceed that shown in figure 2.4 (note though that the education level for an individual remains constant over each individual worklife).

The average wage growth in the economy compounds the productivity growth per efficiency unit of labor shown in figure 2.4 with the effect of increasing educational attainment of the labor force. In addition, there is a small effect arising from changes in the age composition of workers: as we shall see, the experience-wage profile is upward sloping, so an ageing workforce implies somewhat higher average wages. When all these effects are incorporated, the average annual growth rate in the period 2012-2050 is slightly below 5%. We view this forecast as reasonable in light of existing studies. Ge and Yang (2013) estimate an annual 7.7% average wage growth in the period 1992-2007. Concerning future years, we are not aware of independent long-run forecasts of wage growth. However, Citibank forecasts an annual growth rate of GDP per capita of 5% over the period 2010-2050 (Buiter and Rahbari 2011, p.63). If the labor share remained constant, wage growth should remain aligned with GDP growth. In section 2.5.1 we do sensitivity analysis on the future wage growth.

The rate of return on capital is very large in China (see, e.g., Bai et al., 2006). However, these high rates of return appear to have been inaccessible to the government and to the vast majority of workers and retirees. Indeed, in addition to housing and consumer durables, bank deposits are the main asset held by Chinese households in their portfolio. For example, in 2002 more than 68% of households' financial assets were held in terms of bank deposits and bonds, and for the median decile of households this share is 75% (source: Chinese Household Income Project, 2002). Moreover, aggregate household deposits in Chinese banks amounted to 76.6% of GDP in 2009 (source: CSY, 2010). High rates of return on capital do not appear to have been available to the government, either. Its portfolio consists mainly of low-yield bonds denominated in foreign currency and equity in state-owned enterprises, whose rate of return is lower than the rate of return to private firms (see Dollar and Wei, 2007).

Building on Song et al. (2011), the model of section 2.6 provides an explanation – based on large credit market imperfections – for why neither the government nor the workers have access to the high rates of return of private firms. In this section, we simply assume that the annual rate of return for private and government savings is $R = 1.025$. We view a 2.5% annual return for the government savings as realistic. According to the National Council for Social Security Fund, the average share of pension funds invested in stock markets was 19% in 2003-2011.[22] Assuming an average 6% annual return on stock and a 1.75% return on the remaining portfolio yields an average annual return of roughly 2.5%. This is also in line with the return on best-practice Western pension funds. For instance, the Credit Suisse Swiss Pension Fund has achieved a 2.25% annual rate of return between 2000-12. Concerning the return on private savings, a one-year real deposit rate in Chinese banks – the most typical saving instrument of private agents – was 1.75% during 1998-2005 (nominal deposit rate minus CPI inflation). Given that some households have access to savings instruments that yield higher returns, a 2.5% return seems a plausible assumption also for private agents.

Note that our economy is dynamically efficient. Assuming $R < 1.02$ would imply that the rate of return is lower than the long-run growth rate of the economy, implying dynamic

[22] Source: http://www.ssf.gov.cn/xw/xw_gl/201205/t20120509_4619.html.

inefficiency. In such a scenario, there would be no need for a pension reform due to a well-understood mechanism (cfr. Abel et al., 1989).

Consider, finally, preference parameters: the discount factor is set to $\beta = 1.0179$ to capture the large private savings in China. This is slightly higher than the value (1.011) that Hurd (1989) estimated for the United States. As a robustness check, we also consider an alternative economy where β is lower for all people born after 2013 (see section 2.5). In section 2.6 we document that with $\beta = 1.0179$ the model economy matches China's average aggregate saving rate during 2000-2010.

We assume that preferences are represented by the following standard utility function (cfr. equation (2.1)):

$$u(c,h) = \log c - \frac{1}{1+\theta} h^{1+\frac{1}{\theta}},$$

where θ is the Frisch elasticity of labor supply. We set $\theta = 0.5$, in line with standard estimates in labor economics (Keane, 2011). Note that both the social security tax and pensions in old age distort labor supply.

Finally, we obtain the initial distribution of wealth in year 2000 by assuming that all private agents alive in 1992 had zero wealth. We view this as a reasonable assumption, since China's market reforms started in 1992 and most private wealth was accumulated thereafter. Given the 1992 distribution of wealth for workers and retirees, we simulate the model over the 1992-2000 period, assuming an annual wage growth of 5.7%, excluding human capital growth. The distribution of private wealth in 2000 is then obtained endogenously.

2.3.4 The pension system

In this section, we lay out a set of taxes and pension entitlements that replicates the main features of China's system (see Appendix B.3 for a more detailed description of the actual system).

The current Chinese system was originally introduced in 1986 and underwent a major reform in 1997. Before 1986, urban firms (which were almost entirely state owned at that time) were responsible for paying pensions to their former employees. This enterprise-based system became untenable in a market economy where firms can go bankrupt and workers can change jobs. The 1986 reform introduced a defined benefits system whose

administration was assigned to municipalities. The new system came under financial distress, mostly due to firms evading their obligations to pay pension contributions for their workers.

The subsequent 1997 reform reduced the replacement rates for future retirees and tried to enforce social security contributions more strictly. The 1997 system has two tiers (plus a voluntary third tier). The first is a standard transfer-based basic pension system with resource pooling at the provincial level. The second is an individual accounts system. However, as documented by Sin (2005, p.2), "the individual accounts are essentially 'empty accounts' since most of the cash flow surplus has been diverted to supplement the cash flow deficits of the social pooling account." Due to its low capitalization, the system can be viewed as broadly transfer-based, although it permits, as does the US Social Security system, the accumulation of a trust fund to smooth the aging of the population. Since the individual accounts are largely notional, we decided to ignore any distinction between the different pension pillars in our analysis.

We model the pension system as a defined benefits plan, subject to the intertemporal budget constraint, (2.16). Appendix B.3 shows explicitly how the institutional details are mapped into the simple model. In line with the actual Chinese system, pensions are partly indexed to wage growth. We approximate the benefit rule by a linear combination of the average earnings of the beneficiary at the time of retirement and the current wage of workers, with weights 60% and 40%, respectively. More formally, the pension received at period $t+j$ by an agent who worked until period $t+J_W$ (and who became adult in period t) is:[23]

$$b_{t,t+j} = q_{t+J_W} \cdot \left(0.6 \cdot \bar{y}_{t+J_W} + 0.4 \cdot \bar{y}_{t+j-1}\right), \tag{2.14}$$

where $j > J_W$, and q_t denotes the replacement rate in period t and $\bar{y}_t$ is the average pre-tax labor earnings for workers in period t:

$$\bar{y}_t \equiv \frac{w_t \sum_{j=0}^{J_W} N_{t-j,t} \eta_{t-j} \zeta_j h_{t-j,t}}{\sum_{j=0}^{J_W} N_{t-j,t}}. \tag{2.15}$$

[23] Alternatively, the law of motion of pension benefits can be expressed as $b_{t,t+j} = b_{t+J_W+1}\left(0.6 + 0.4 \times \left(\bar{y}_{t+j-1}/\bar{y}_{t+J_W}\right)\right)$.

In line with the 1997 reform (see, e.g., Sin, 2005), we assume that pensioners retiring before 1997 continued to earn a 78% replacement rate throughout their retirement. Moreover, those retiring between 1997 and 2011 are entitled to a 60% replacement ratio.

We assume a constant social security tax (τ) equal to 20%, in line with the empirical evidence.[24]

The current pension system of China covers only a fraction of the urban workers. The coverage rate has grown from 45% in 2001 to 60% in 2011 (see *China Statistical Yearbook* 2012). In the baseline model, we therefore assume a constant coverage rate of 60%. Workers who are not covered neither pay the social security tax nor do they receive pensions.

The coverage rate of migrant workers is a key issue. Since we do not have direct information about their coverage, we decided to simply assume that rural immigrants get the same coverage rate as urban workers. This seems a reasonable compromise between two considerations. On the one hand, the coverage of migrant workers (especially low-skill non-hukou workers) is lower than that of non-migrant urban residents; on the other hand, the total coverage has been growing since 1997.[25]

2.3.5 The government budget constraint

The pension system is said to be *financially balanced* if, given an initial government wealth A_0. the government intertemporal budget constraint holds, i.e.,

$$\sum_{t=0}^{\infty} R^{-t} \left(\sum_{j=J_W+1}^{J} N_{t-j,t} b_{t-j,t} - \tau_t \sum_{j=0}^{J_W} N_{t-j,t}\ \zeta_j \eta_{t-j} w_t\ h_{t-j,t} \right) \leq A_0, \tag{2.16}$$

where $N_{t-j,t}$ is the number (measure) of agents in period t who became active in period $t-j$. Equation (2.16) is the analogue of equation (2.2) in the two-period model of section 2.2.

[24] The statutory contribution rate including both basic pensions and individual accounts is 28%. However, there is evidence that a significant share of the contributions is evaded, even for workers who formally participated in the system. See the webpage appendix for details.

[25] According to a recent document issued by the National Population and Family Planning Commission, 28% of migrant workers are covered by the pension system (Table 5-1, 2010 Compilation of Research Findings on the National Floating Population).

We set the initial government wealth, A_0, equal to 71% of GDP in 2000. As we explain below in Section 2.6.1, this level is consistent with the observed foreign surplus in year 2000, given the (endogenous) level of private wealth in the same year. The entire government wealth is assumed to enter the budget of the pension system.

2.3.6 The benchmark reform

Under our calibration of the model, the current pension system is not balanced. In other words, the intertemporal budget constraint, (2.16), would not be satisfied if the current rules were to remain in place forever. For the intertemporal budget constraint to hold, it is necessary either to reduce pension benefits or to increase contributions.

We construct a benchmark pension system to which we compare alternative reforms. To ensure that this system is financially viable, we assume that (i) the existing rules apply for all workers who are already retired by 2013; (ii) the social security tax remains constant $\tau = 20\%$ for all cohorts; (iii) for workers retiring in 2013 or later, the replacement rate is amended and set permanently to a new level q which is the highest constant level consistent the intertemporal budget constraint, (2.16). All households are assumed to anticipate that the benchmark reform will take place in 2013. We refer to such a scenario as the *benchmark reform*.[26]

The benchmark reform entails a large reduction in the replacement rate, from 60% to 40%. Namely, pensions must be cut by a third in order for the system to be financially sustainable. Such an adjustment is consistent with the existing estimates of the World Bank (see Sin, 2005, p.30). Alternatively, if one were to keep the replacement ratio constant at the initial 60% and to increase taxes permanently so as to satisfy (2.16), then τ should increase from 20% to 30.2% as of year 2013.

Figure 2.5 shows the evolution of the replacement rate by cohort under the benchmark reform (panel (a), dashed line). The replacement rate is 78% until 1997 and then falls to 60%. Under the benchmark reform, it falls further to 40% in 2013, remaining constant

[26]It is natural to impose financial balance on the assumed benchmark policy. It would not make sense to compare an unbalanced system with reforms that satisfy the intertemporal budget constraint of the government. For this reason, we cannot consider as a benchmark a system based on keeping forever the current rules.

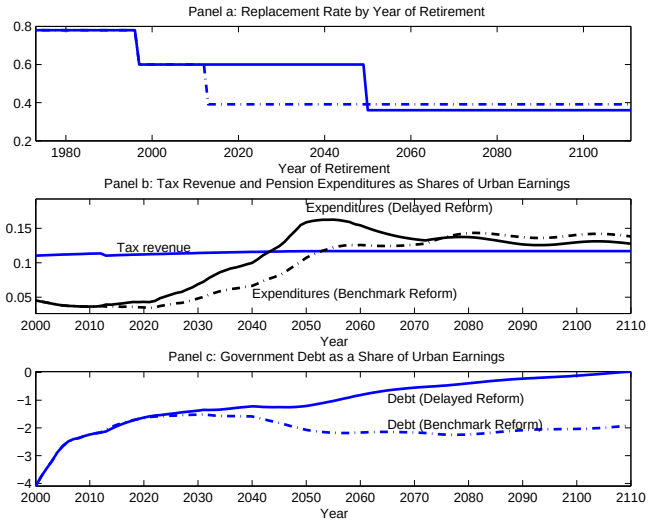

Figure 2.5: Delayed reform

Panel (a) shows the replacement rate q_t for the benchmark reform (dashed line) versus the case when the reform is delayed until 2050. Panel (b) shows tax revenue and expenditures, expressed as a share of aggregate urban labor income (benchmark reform is dashed and the delay-until-2050 is solid). Panel (c) shows the evolution of government debt, expressed as a share of aggregate urban labor income (benchmark reform is dashed and the delay-until-2050 is solid). Negative values indicate a surplus.

thereafter. Panel (b) (dashed line) shows that such a reform implies that the pension system runs a surplus until 2051. The government builds up a government trust fund amounting to 261% of urban labor earnings by 2080 (panel (c), dashed line). The interests earned by the trust fund are used to finance the pension system deficit after 2051.[27]

2.4 Alternative pension reforms

The theoretical analysis of section 2.2 shows that a social planner with a discount factor no higher than $(1+g)/R$ (where, recall, g is the long run growth rate, and not the transitional wage growth in an emerging economy) would want to redistribute in favor of the poorer earlier generations. The benchmark reform, to the opposite, reduces current pension payments drastically in order to guarantee the financial sustainability of the pension in the long run.

In this section, we consider a set of alternative reforms that are also financially sustainable, but distribute the costs and benefits of the adjustment in a different way from the benchmark reform. We first consider a set of theoretically motivated reforms along the lines of what we studied in section 2.2. Then, we consider a set of more "policy-driven" reforms that would alter in a less extreme fashion the existing rules, or that corresponds to actual reforms that have been on the agenda of the policy debate. Each alternative policy reform is introduced as a “surprise”. Namely, agents expect the benchmark reform, but when 2013 arrives, unexpectedly, they learn that a different reform will take place. Subsequently, perfect foresight is assumed. This assumption is not essential. In section 2.5.5, we show that the main results are robust to assuming, alternatively, that all reforms are perfectly anticipated in year 2000.[28]

[27] Note that in panel c the government net wealth (i.e., minus the debt) is falling sharply between 2000 and 2020 when expressed as a share of urban earnings, even though the government is running a surplus. This is because urban earnings are rising very rapidly due to both high wage growth and growth in the number of urban workers.

[28] Conceptually, assuming that alternative reforms come as a surprise has the advantage that, when we derive the wage path from a model with endogenous capital accumulation (section 2.6), all macroeconomic trends which are the target of the calibration are identical until 2013. Moreover, from a policy perspective, it seems natural to study the effect of policy changes happening today without retrospectively affecting earlier expectations.

2.4.1 The welfare criterion

Since the main goal of our analysis is to quantify the welfare implications of different reforms, we first introduce a welfare criterion analogous to that used in the theoretical analysis of section 2.2. To this end, we measure, for each cohort, the equivalent consumption variation of each alternative reform relative to the benchmark reform. Namely, we calculate what (percentage) change in lifetime consumption would make agents in each cohort indifferent between the benchmark and the alternative reform.[29] We then aggregate the welfare effects of different cohorts by means of a utilitarian social welfare function, where the weight of the future generation decays geometrically with a constant factor ϕ as in section 2.2. More formally, the planner's welfare function (evaluated in year 2013) is given by

$$U = \sum_{t=1935}^{\infty} \phi^t N_{t,t} \sum_{j=0}^{J} \beta^j u\left(c_{t,t+j}, h_{t,t+j}\right). \tag{2.17}$$

Then, the equivalent variation is given by the value ω solving

$$\sum_{t=1935}^{\infty} \phi^t N_{t,t} \sum_{j=0}^{J} \beta^j u\left((1+\omega)\, c_{t,t+j}^{BENCH}, h_{t,t+j}^{BENCH}\right) = \sum_{t=1923}^{\infty} \phi^t N_{t,t} \sum_{j=0}^{J} \beta^j u\left(c_{t,t+j}^{*}, h_{t,t+j}^{*}\right), \tag{2.18}$$

where superscripts *BENCH* stand for the allocation in the benchmark reform and asterisks stand for the allocation in the alternative reform.[30]

The planner experiences a welfare gain (loss) from the alternative allocation whenever $\omega > 0$ ($\omega < 0$). We shall consider two particular values of the intergenerational discount factor, ϕ. First, $\phi = (1+g)/R$, which is the benchmark discount factor discussed in section 2.2 (see Proposition 2.1 and its Corollary) corresponding to a planner would not want to engage in any intergenerational redistribution in steady state. Since in our calibration $R = 1.025$ and $g = 0.02$, such a planner has an annual discount rate of 0.5%, a small number relative to standard calibrations.[31] For this reason, we label the planner with $\phi = (1+g)/R$

[29] Note that we measure welfare effects relative to increases in *lifetime* consumption even for people who are alive in 2012. This approach makes it easier to compare welfare effects across generations.

[30] Note that we sum over agents alive or yet unborn in 2012. The oldest person alive became an adult in 1935, which is why the summations over cohorts indexed by t start from 1935.

[31] Most macroeconomic studies assume discount rates in the range 3-5%. In the debate on global warming, Nordhaus suggests a 3% discount rate. Stern argues that this is ethically indefensible, and proposes to apply

as the *low-discount planner*. As a robustness, following Nordhaus (2007), we consider the case of $\phi = R^{-1}$, namely, the planner discounts future utilities at the market interest rate. We label such a planner as the *high-discount planner.* Relative to the low-discount benchmark, the high-discount planner will demand more intergenerational redistribution in favor of the earlier generations.

2.4.2 Theory-driven reforms

In this section, we compute the pension systems that implement the first- and second-best choices of a low-discount planner, and compare it with the benchmark reform. We consider, additionally, a more restrictive environment in which the planner cannot increase the generosity of the pension system relative to the existing rules, namely, future replacement rates cannot exceed 60% (whereas the existing rules apply for the agents already retired in 2013).

The two panels of figure 2.6 show, respectively, the sequence of cohort-specific replacement rates in each of the three alternative reforms (upper panel), and the consumption equivalent welfare gain for each cohort relative to the benchmark reform (lower panel). The panels display only generations retiring after 2000.[32]

Consider the first best. The replacement rate is 235% for the cohort retiring in 2013. Thereafter, it falls roughly linearly with the retirement date until it reaches -22.2% in 2072. There are huge welfare gains for the transition generations – exceeding 100% for those retiring between 2013 and 2032. The welfare gains fall over time and converge to -8.5% for the cohort retiring after 2073. All generations retiring before 2061 gain from the reform. The welfare gain accruing to the low-discount planner is 3.8% of consumption. In the case of the high-discount planner the gain is a staggering 41.9%.

The second best (with non-negative benefits) yields a similar picture, although it delivers slightly lower replacement rates for the transition generations, reaching zero for cohorts

a 0.1% discount rate. Such a low discount rate would be largely inconsistent with the observation that governments typically run debt rather accumulate surpluses, despite the fact that the interest rate is on average significantly above 0.1%. In this paper, we consider discount rates ranging between 0.5% to 2.5%, which we regard as a conservative criterion.

[32]The efficient scheme involves large transfers to the generations already retired. For instance, those retiring in 1990 receive a replacement rate equal to 738% in the first best and to 698% in the second best.

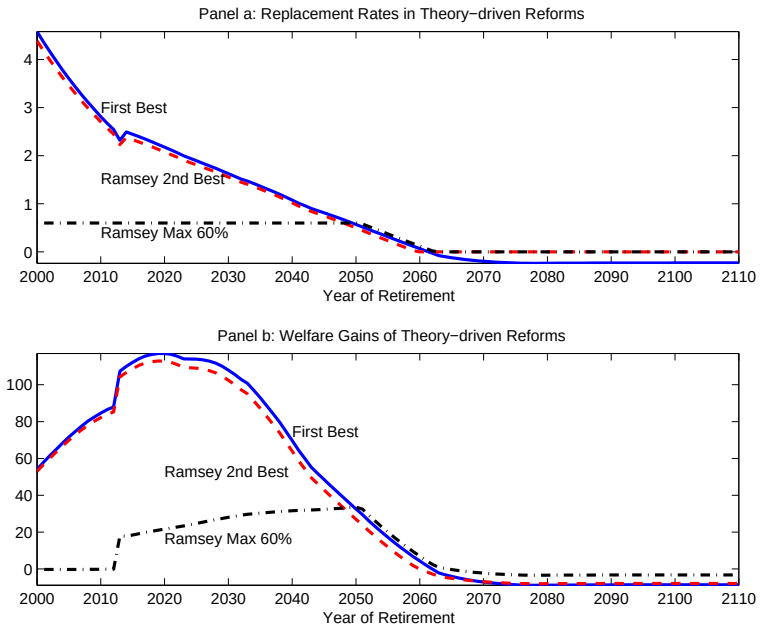

Figure 2.6: Theory driven reforms

Panel (a) plots the sequence of cohort-specific replacement rates in the first best reform (blue solid line), second-best Ramsey reform with non-negative pensions (red dashed line), and Ramsey reform where future replacement rates are bounded between zero and 60% (black dash-dotted line). Panel (b) plots the corresponding consumption equivalent welfare gains for each cohort.

retiring after 2060. Taxes are zero for cohorts retiring before 2059, implying that the system builds up a debt that is financed by taxes on future generations. In steady state, the tax rate reaches 9.7%. The welfare gain to the low-discount planner amounts to 3.6% of consumption.[33]

Finally, consider the constrained Ramsey allocation where the replacement rate must stay between 0 and 60%. In this case, the replacement rate is exactly 60% for all cohorts retiring until 2050. The replacement rate falls and reaches zero in 2063. The steady-state taxes are lower (5.2%), because the pension system is less generous with the transition generation and does not build up such a large debt as in the previous case. The welfare gain to the low-discount planner is now substantially lower but still significant, being equal to 2% of consumption.

In conclusion, the quantitative normative analysis of this section has shown that even a planner with a very high weight on future generations would use the pension system to implement a radical intergenerational redistribution *in spite* of the averse demographics.

2.4.3 Policy-driven reforms

The benchmark reform achieves financial balance through a draconian permanent reduction in pension entitlements for all agents retiring after 2012. The analysis in section 2.4.2 shows that such adjustment puts too large a burden on current generations relative to the normative benchmark.

The optimal pension policies discussed above are informative about how to improve on the benchmark reform, but arguably difficult to implement For instance, much of the current debate focuses on whether reforms reducing the generosity of the system are urgent or can be postponed, and on whether China should adopt rules that nudge the system in a more funded direction.

[33] We computed the first- and second-best (and the corresponding benchmark reform) under the alternative assumption that $A_{2013} = 0$. The results are similar. The welfare gain of the first-best increases from 3.75% to 3.79%, while the second-best delivers smaller gains (3.67% vs. 3.64%). The planner delivers positive pensions till 2058, and the steady state tax rate reaches 10.2%.

In this section, we consider a set of alternative sustainable reforms that speak more directly to the policy debate, and that would alter in a less extreme way the existing rules. We consider three types of reforms:

1. *Delayed reform*: we assume that the current rules are kept in place until period T (where $T > 2013$), in the sense that the current replacement rate ($q_t = 60\%$) applies for those who retire until period T, and taxes remain at 20%. Thereafter, the replacement rates are adjusted permanently so as to satisfy (2.16). Note that, since the current system is not financially balanced, a delay requires a larger cut in replacement rates after T. T is chosen optimally so as to maximize the planner's welfare. This reform entails a key aspect of the optimal policy: the replacement rate is decreasing over time, providing intergenerational distribution from the future richer generations to the current poorer transition generation.

2. *Fully-funded (FF) reform*: we replace the current transfer-based system with a mandatory saving-based scheme in 2013. In the FF reform scenario, defined benefit transfers are abolished in 2013. However, the government does not default on its outstanding liabilities: those who are already retired receive a lump-sum transfer equal to the present value of the benefits they would have received under the benchmark reform. Moreover, those still working in 2013 are compensated for their accumulated pension rights, scaled by the number of years they have contributed to the system. This reform entails an aspect of the optimal policy: it reduces the distortion caused by the social security tax, although it does not provide any intergenerational redistribution.

3. *Pay-as-you-go (PAYGO) reform*: we impose an annual balanced budget requirement to the pension system, keeping the social security tax at 20%. The benefit rate is endogenously determined by the tax revenue (which is, in turn, affected by the demographic structure and endogenous labor supply). Given the demographic transition and the initially high wage growth, this reform yields high pensions to the earlier generations, and low pensions to the future ones – in line with the optimal policy.

Finally, we consider two reforms that extend the coverage of the pension system to rural workers. The *moderate* rural reform scenario offers a 20% replacement rate to rural retirees

financed by a 6% social security tax on rural workers. Such a rural pension is similar to a scheme started recently by the government on a limited scale (see Appendix B.3 for details). The *radical* rural reform scenario introduces a universal pension system with the same benefits and taxes in rural and urban areas.

2.4.3.1 Delayed reform

We start by computing the optimal delay of the benefit cut. The optimal T for the low-discount planner turns out to be 2050. Namely, the current replacement rate continues to apply for all workers starting their employment before 2011, and the new lower replacement rate applies to worker starting their employment earliest 2012. This means that lower pensions will start being paid in 2050, and by 2090 all retirees will earn the new lower replacement rate.

Due to the delay, the fund accumulates initially a lower surplus, forcing a larger reduction of the replacement rate after 2050. Thus, relative to the benchmark reform, the delay shifts the burden of the adjustment from the current (poorer) generations to (richer) future generations.

Figure 2.5 describes the positive effects of delaying the reform until 2050. Panel (a) shows that the post-reform replacement rate now falls to 37%, which is only 3 percentage points lower than the replacement rate granted by the benchmark reform. Panel (b) shows that the pension expenditure is higher than in the benchmark reform until 2075. Moreover, already in 2043 the system runs a deficit.

Figure 2.7 shows the welfare gains of four reforms relative to the benchmark, broken down by the year of retirement of each cohort. Consider the delayed reform experiment: There are large gains for agents retiring between 2013 and 2049, on average over 15.3% of their lifetime consumption. The main reason is that delaying the reform enables the transition generation to share the gains from high wage growth after 2013, to which pension payments are (partially) indexed. The welfare gain declines over the year of cohort retirement, since wage growth slows down. All generations retiring after 2050 lose, although their welfare losses are quantitatively small, being less than 1.6% of their lifetime consumption. Relative to the first best, the delayed reform implies too little intergenerational redistribution from future to current generations. Moreover, it entails labor supply

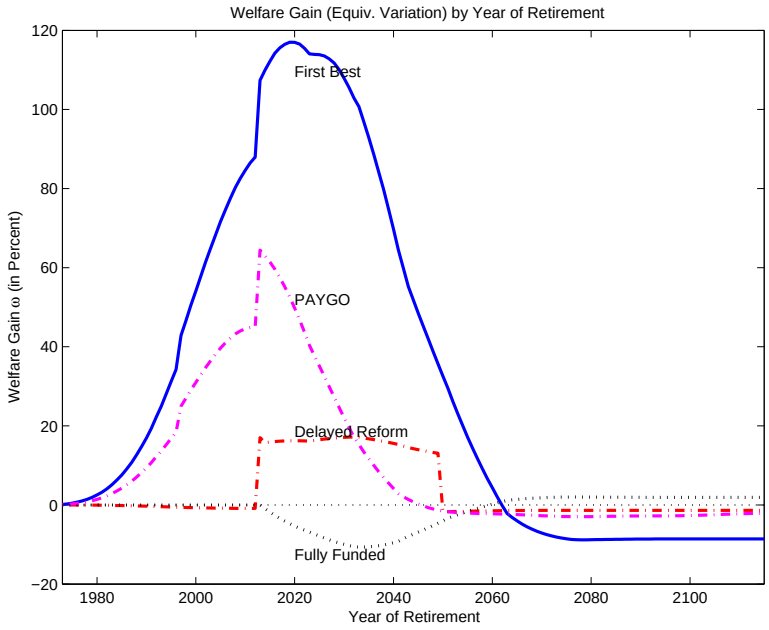

Figure 2.7: Welfare gain

The figure shows welfare gains of the policy-driven alternative reforms relative to the benchmark reform for each cohort. For comparison, the welfare effects of the first-best policy is also plotted. The gains (ω) are expressed as percentage increases in consumption (see eq. 2.18).

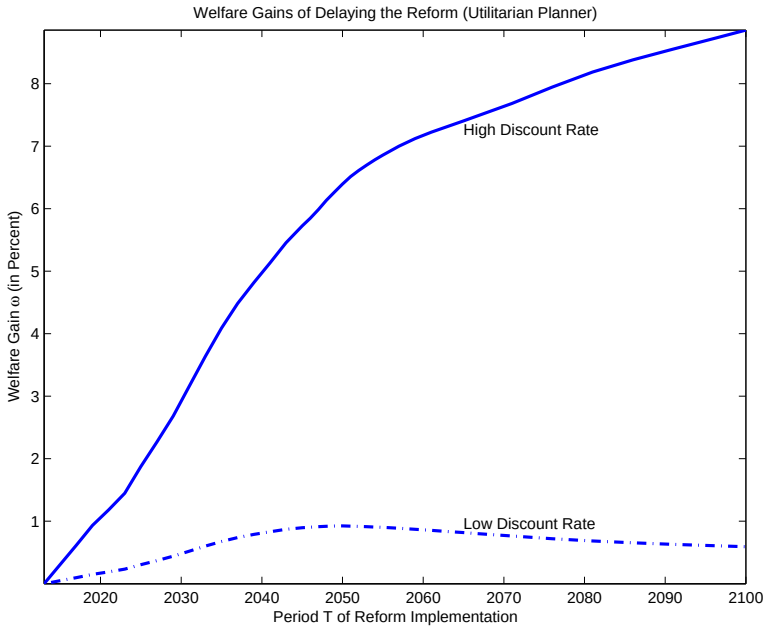

Figure 2.8: Welfare gain of delaying the reform

The figure shows the consumption equivalent gain/loss accruing to a high-discount planner (solid line) and to a low-discount planner (dashed line) of delaying the reform until time T relative to the benchmark reform. When $\omega > 0$, the planner strictly prefers the delayed reform over the benchmark reform.

distortions that are absent in the first best. Yet, the low discount planner enjoys a 0.9% welfare gain, corresponding to roughly one quarter of the potential gain in the first best, and half of the welfare gain obtained in the planning allocation subject to the constraint that the replacement must lie between zero and 60%.

Figure 2.8 shows the welfare gains/losses of delaying the reform until year T. The figure displays two curves: in the upper curve, we have the consumption equivalent variation of the high-discount planner, while in the lower curve we have that of the low-discount planner. As discussed above, it is optimal for the low discount planner to delay the reform until 2050. The same delay would yield a much larger welfare gain (6.2%) for the high-discount planner whose utility is increasing in the entire range plotted by the figure.

2.4.3.2 Fully Funded Reform

Consider, next, switching to an FF system, i.e., a pure contribution-based pension system featuring no intergenerational transfers, where agents are forced to save for their old age in a fund that has access to the same rate of return as that of private savers. As long as agents are rational and have time-consistent preferences, and mandatory savings do not exceed the savings that agents would make privately in the absence of a pension system, an FF system is equivalent to no pension system.[34] As discussed above, the government does not default on existing claims: All workers and retirees who have contributed to the pension system are refunded the present value of the pension rights they have accumulated.[35] Since the social security tax is abolished, the existing liabilities are financed by issuing government debt. This debt is rolled over and serviced by a constant labor income tax (implying that the outstanding debt level can fluctuate over time). This scheme is similar to that adopted in the 1981 pension reform of Chile.

Figure 2.9 shows the outcome of this reform. The old system is terminated in 2013, but people with accumulated pension rights are compensated as discussed above. To finance such a pension buy out scheme, government debt must increase to over 89.5% of total labor earnings in 2012. A permanent 0.3% annual tax is needed to service the debt. The government debt first declines as a share of total labor earnings due to high wage growth in that period, and then stabilizes at a level about 30% of labor earnings around 2040. Future generations live in a low-tax society with no intergenerational transfers.

As shown in figure 2.7, the distributional effects are opposite to those of the delayed reforms. The cohorts retiring between 2013 and 2058 are harmed by the FF reform relative to the benchmark. There is no effect on earlier generations, since those are fully compensated by assumption. The losses are also modest for cohorts retiring soon after 2013, since these have earned almost full pension rights by 2013. However, the losses increase for later cohorts and become as large as 11% for those retiring in 2030-35. For such cohorts, the

[34] Bohn (2010) shows that such equivalence breaks down in the presence of political or financial constraints. These aspects are ignored in our paper.

[35] In particular, people who have already retired are given an asset worth the present value of the pensions according to the old rules. Since there are perfect annuity markets, this is equivalent to the pre-reform scenario for those agents. People who are still working and have contributed to the system are compensated in proportion to the number of years of contributions.

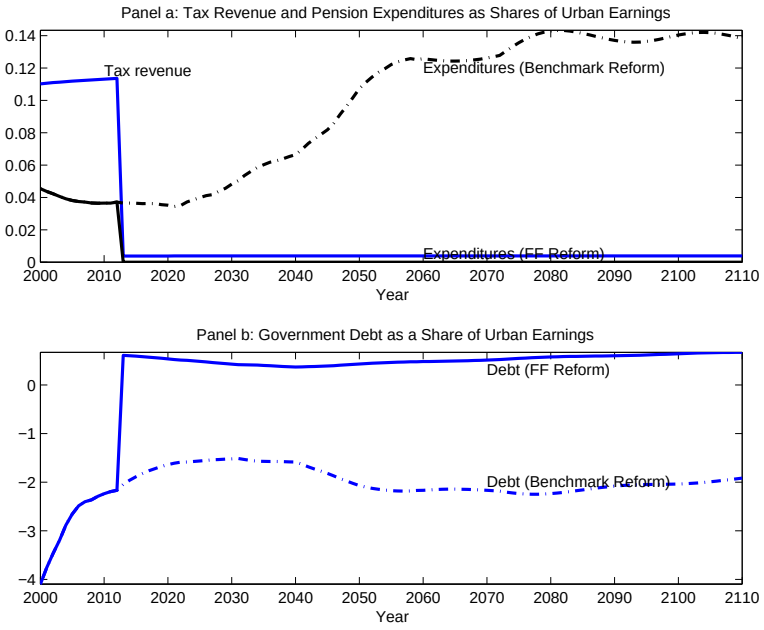

Figure 2.9: Fully funded reform

The figure shows outcomes for the fully funded reform (solid lines) versus the benchmark reform (dashed lines). Panel (a) shows the replacement rates. Panel (b) shows the government debt as a share of aggregate urban labor income.

system based on intergenerational transfer is attractive, since wage growth is high during their retirement age (implying fast-growing pensions), whereas the returns on savings are low. Losses fade away for cohorts retiring after 2050 and turn into gains for those retiring after 2059. The fact that generations retiring sufficiently far in the future gain is guaranteed by the assumption that the economy is dynamically efficient. However, the long-run gains are modest.

The FF reform yields a 0.2% consumption equivalent gain for the low-discount planner. This small gain arises from two opposite effects: on the one hand, the FF reform reduces the labor supply distortion, due to the lower taxes; on the other hand, it does worse than the benchmark reform in terms of the intergenerational redistribution desired by the planner. As the high-discount planner values intergenerational redistribution more than does the low-discount planner, the former strictly prefers the benchmark over the FF reform, with a consumption equivalent discounted loss of 3.3%.

2.4.3.3 Pay-as-you-go reform

The delayed reform experiment was restricted by design to yield a two-tier replacement rate (pre- and post-reform) with a maximum replacement rate of 60% for the generations before the reform. In contrast, as shown above, the optimal policy feature a declining benefit sequence with very high replacement rates for the initial generations (particularly, those already retired). In an aging economy, a pure PAYGO system would precisely yield a smooth decline in replacement rates. However, relative to the optimal policy, a PAYGO entails tax distortions that the planner, as we showed, dislikes.

In this section, we consider the effect of switching to a PAYGO. We maintain the contribution rate fixed at $\tau = 20\%$ and assume that the benefits equal the total contributions in each year. Therefore, the pension benefits b_t in period t are endogenously determined by

the following formula:[36]

$$b_t = \frac{\tau \sum_{j=0}^{J_W} N_{t-j,t}\ \zeta_j \eta_{t-j} w_t\ h_{t-j,t}}{\sum_{j=J_W+1}^{J} N_{t-j,t}}.$$

Figure 2.10 shows the outcome of this reform. Panel (a) reports the pension benefits as a fraction of the average earnings by year. Note that this notion of replacement rate is different from that used in the previous experiments (panel a of figures 2.5, and 2.9); there the replacement rate was cohort specific and was computed according to equation (2.14) by the year of retirement of each cohort. Until 2050, the PAYGO reform implies larger average pensions than under the benchmark reform.

Panel (b) shows the lifetime pension as a share of the average wage in the year of retirement, by cohort. This is also larger than in the benchmark reform until the cohort retiring in 2045. We should note that, contrary to the previous experiments which were neutral vis-à-vis cohorts retiring before 2013, here even earlier cohorts benefit from the PAYGO reform, since the favorable demographic balance yields higher pensions than what they were promised. This can be seen in panel (b) of figure 2.10 and figure 2.7. Welfare gains are very pronounced for all cohorts retiring before 2045, especially so for those retiring right after 2013, who would suffer a significant pension cut in the benchmark reform. These cohorts retire in times when the old-age dependency ratio is still very low, so benefits are large. Generations retiring after 2046, instead, lose.

Due to the strong redistribution in favor of poorer early generations, in spite of the tax distortion, the utilitarian welfare is significantly higher under the PAYGO reform than in the benchmark reform, for both a high- and low-discount planner. The consumption equivalent gains relative to the benchmark reform are, respectively, 12.9% and 1.7% for urban workers. These gains are larger than under all alternative reforms (including delayed and FF reform).

[36]Note that the pension system has accumulated some wealth before 2012. We assume that this wealth is rebated to the workers in a similar fashion as the implicit burden of debt was shared in the fully funded experiment. In particular, the government introduces a permanent reduction δ in the labor income tax, in such a way that the present value of this tax subsidy equals the 2012 accumulated pension funds. In our calibration, we obtain $\delta = 0.59\%$.

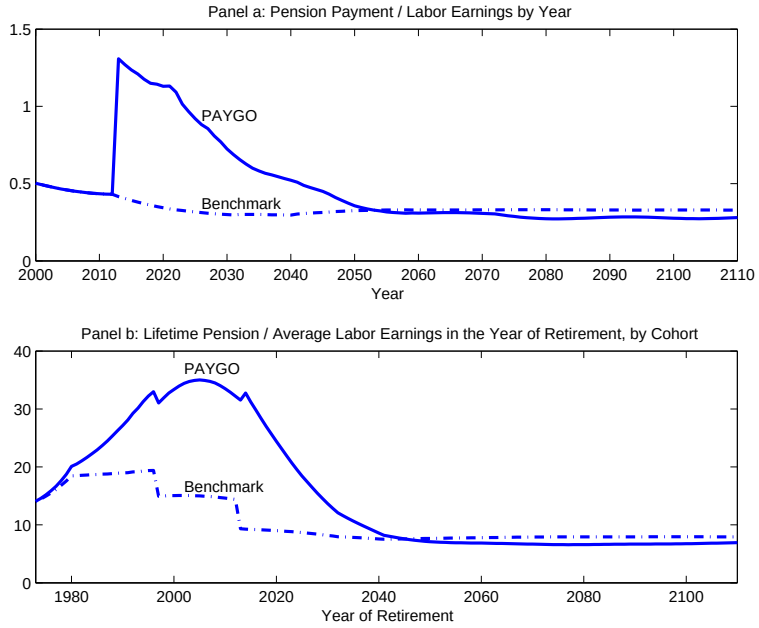

Figure 2.10: PAYGO reform

Panel (a) shows the average pension payments in year t as a share of average wages in year t for the PAYGO (solid) and the benchmark reform (dashed line). Panel (b) shows the ratio of the lifetime pensions (discounted to the year of retirement) to the average labor earnings just before retirement for each cohort.

2.4.4 Other reforms

In this section, we consider other reforms that have been discussed in the context of China: increasing retirement age and increasing the coverage of the pension system to rural workers.

2.4.4.1 Increasing retirement age

An alternative to reducing pension benefits would be to increase the retirement age. Our model allows us to calculate the increase in retirement age that would be required to balance the intertemporal budget, (2.16), given the current social security tax and replacement rate. We find such an increase to be equal to approximately six years (i.e., retirement age would have to increase from 60 to 66 years without any reduction in employment). This shows that a draconian reduction in pension entitlements may not be necessary if the retirement age can be increased. Since our model abstracts from an endogenous choice of retirement, we do not emphasize the welfare effects of policies affecting retirement age (there would obviously be a large welfare gain if the retirement age is increased exogenously).

2.4.4.2 Rural Pension

The vast majority of people living in rural areas are not covered by the current Chinese pension. In accordance with this fact, we have so far maintained the assumption that only urban workers are part of the pension system. In this section, we consider extending the system to rural workers.

Although a rural and an urban pension system could in principle be separate programs, we assume that there is a consolidated intertemporal budget constraint, namely, the government can transfer funds across the rural and urban budget. This is consistent with the observation that the modest rural pension system that China is currently introducing is heavily underfunded (see Appendix B.3), suggesting that the government implicitly anticipates a resource transfer from urban to rural areas. The modified consolidated government

budget constraint then becomes

$$A_0+\sum_{t=0}^{\infty}R^{-t}\left(\sum_{j=0}^{J_W}\zeta_j\left[\tau_t N_{t-j,t}\,w_t\,h_{t-j,t}+\tau_t^r N_{t-j,t}^r\,w_t^r\,h_{t-j,t}^r\right]-\sum_{j=J_W+1}^{J}\left[N_{t-j,t}b_{t-j,t}+N_{t-j,t}^r b_{t-j,t}^r\right]\right)\geq 0, \tag{2.19}$$

where superscripts r denote variables pertaining to the rural areas, whereas urban variables are defined, as above, without any superscript.

We assume the rural wage rate to be 54% of the urban wage in 2000, consistent with the empirical evidence from the China Health and Nutrition Survey. The annual rural wage growth is assumed to be on average 4.1% between 2000-2024, and 2% thereafter. This is consistent with the prediction of the general equilibrium model outlined in section 2.6.

We consider two experiments. In the first (*low-scale reform*), we introduce a rural pension system with rules that are different from those applying to urban areas in 2013. This experiment mimics the rules of the new old-age programs that the Chinese government is currently introducing for rural areas (see Appendix B.3). Based on the current policies, we set the rural replacement rate (q_t^r) and contribution rate (τ_t^r) to 20% and 6%, respectively. These rates are assumed to remain constant forever. Moreover, we assume that all rural inhabitants older than retirement age in 2013 are eligible for this pension. Introducing such a scheme in 2013 would worsen the fiscal imbalance. Restoring the fiscal balance through a reform in 2013 requires that the replacement rate of urban workers be cut to $q_t = 38.7\%$, 1.3 percentage points lower than in the benchmark reform without rural pensions. Hence, the rural pension implies a net transfer from urban to rural inhabitants.

A low-discount planner who only cares for urban households participating in the pension system would incur a welfare loss of less than 0.7% from expanding the pension system to rural inhabitants. In contrast, a low-discount planner who only cares for rural households would incur a welfare gain of 7.5%. When weighting rural and urban households by their respective population shares, one obtains an aggregate welfare gain of 1% relative to the benchmark reform.

The second experiment (*drastic reform*) consists of turning the Chinese pension system into a universal system, pooling all Chinese workers and retirees – in both rural and urban areas – into a system with common rules. As of 2013, all workers contribute 20% of

their wage. In addition, the system bails out all workers who did not contribute to the system in the past. Namely, all workers are paid benefits according to the new rule even though they had not made any contribution in the past. Although rural and urban retirees have the same replacement rate, pension benefits are proportional to the group-specific wages (i.e., rural [urban] wages for rural [urban] workers). As in the benchmark reform above, the replacement rate is adjusted in 2013 so as to satisfy the intertemporal budget constraint of the universal pension system. Although we ignore issues with the political and administrative feasibility of such a radical reform, this experiment provides us with an interesting upper bound of the effect of a universal system.

The additional fiscal imbalance from turning the system into a universal one is small: the replacement rate must be reduced to $q_t = 38.5\%$ from 2013 onward, relative to 40% in the benchmark reform. The welfare loss for urban workers participating in the system is very limited (only marginally higher than in the low-scale reform). In contrast, there are sizable welfare gains for rural and urban workers not participating in the system (an average 4.5% and 1%, respectively, if evaluated by a low-discount planner).

To understand why this reform can give so large gains with such a modest additional fiscal burden, it is important to emphasize that (i) the earnings of rural workers are on average much lower than those of urban workers; and (ii) the rural population is declining rapidly over time. Both factors make pension transfers to the rural sector relatively inexpensive. It is important to note that our calculations ignore any cost of administering and enforcing the system. In particular, the benefit would decrease if the enforcement of the social security tax in rural areas proves to be more difficult than in urban areas.

2.5 Sensitivity analysis

In this section, we study how the main results of the previous section depend on structural features of the model economy: wage growth, population dynamics, and interest rate. For simplicity, we focus on the urban pension system (no payments to rural workers). We refer to the calibration of the model used in the previous section as the *baseline economy*. Table 2.1 summarizes the results discussed throughout this section. Each row reports the welfare

	Delayed until 2050		Delayeduntil2100		FullyFunded		PAYGO	
discount	high	low	high	low	high	low	high	low
Baseline parameterization	6.2%	0.9%	8.5%	0.6%	-3.3%	0.2%	12.9%	1.7%
Slow wage convergence	6.2%	1.0%	8.7%	0.7%	-3.6%	0.1%	12.6%	1.8%
Low wage growth	3.3%	-0.1%	5.6%	-0.4%	-0.6%	0.8%	4.9%	-0.1%
Low fertility	7.8%	3.0%	10.3%	0.6%	-2.4%	-0.5%	15.7%	5.0%
Slow migration	6.2%	0.9%	8.4%	0.5%	-3.3%	0.2%	11.9%	1.6%
High interest rate	2.7%	-0.1%	4.0%	-0.4%	0.5%	0.5%	8.7%	-0.1%
Perfect foresight	6.2%	0.9%	8.5%	0.6%	-3.3%	0.2%	14.7%	2.0%

Table 2.1: Welfare gains/losses

Welfare gains/losses under alternative reforms. The table summarizes the welfare effects (measured in terms of compensated variation in consumption for the high- and low-discount rate planners, respectively) of alternative pension reforms relative to the benchmark 2013 reform.

effects of different reforms accruing to the high- and low-discount planner relative to a particular environment.

2.5.1 Lower wage growth

In the analysis above, Chinese wages grow fast over the next twenty-five years, and converge to 57% of the US level by 2039. Thereafter, the gap remains constant. In the theoretical model of section 2.2 we established that fast convergence followed by a growth slowdown is the key driver of intergenerational redistribution. In this section, we first consider a model where there is no convergence after 2013. Our analysis verifies that, in line with the results of section 2.2, the model yields "conventional" predictions in this case: the low-discount planner would choose no intergenerational redistribution. Thus, an FF reform is preferred to a PAYGO reform since it reduces tax distortions.

We also study the role of the speed of the transition – this is especially important for China since its future speed of convergence is uncertain. To this aim, we assume that wages converge more slowly to the same long-run gap (57% of the US level) as in the baseline scenario. The main results of the baseline economy hold up: delaying reform is good, and a PAYGO reform dominates an FF reform.

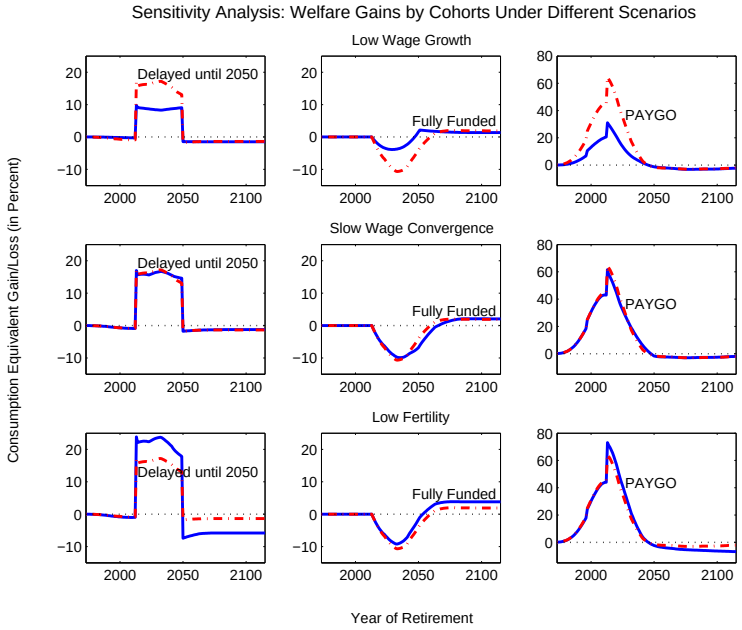

Figure 2.11: Alternative scenarios

The figure shows consumption equivalent gains/losses accruing to different cohorts in three alternative scenarios. The top panels refer to the slow wage converge scenario of section 2.5.1.2. The middle panels refer to the low wage growth (no convergence) scenario of section 2.5.1.1. The bottom panels refer to the low fertility scenario of section 2.5.2. In each panel, the dashed lines refer to the welfare gains under the benchmark calibration (see section 2.4). The left-hand panels show the consumption equivalent gains/losses associated with delaying the reform until 2050 (solid lines). The center panels show the consumption equivalent gains/losses associated with a fully funded reform (solid lines). The right-hand panels show the consumption equivalent gains/losses associated with a PAYGO reform (solid lines).

2.5.1.1 Scenario 1: Low wage growth (no convergence)

In this scenario, we assume wage growth to be constant and equal to 2% after 2013. In this case, the benchmark reform implies a replacement rate of 40.3%.[37]

The welfare effects of the alternative reforms are displayed in the top row of figure 2.11. In general, the welfare gains of the earlier generations relative to the benchmark 2013 reform are significantly smaller than in the baseline wage growth economy. For instance, if

[37]Note that in the low wage growth economy, the present value of the pension payments is lower than in the baseline economy, since pensions are partially indexed to the wage growth. Thus, pensions are actually lower, in spite of the slightly higher replacement rate.

the reform is delayed until 2050 (yielding a replacement rate of 36.9%) the cohorts retiring between 2013 and 2049 experience a welfare gain ranging between 8.2% and 9.6%. The cost imposed on the future generations remains similar in magnitude to that of the baseline economy. For the low-discount planner, there is a tiny loss from delaying. The high-discount planner continues to enjoy a positive welfare gain (3.3%), albeit is significantly lower than in the baseline economy. This is not surprising, since the high-discount planner wants a declining replacement rate sequence even in steady state (see Proposition 2.1).

As in the baseline case, the FF alternative reform harms earlier cohorts, whereas it benefits all cohorts retiring after 2046. However, the relative losses of the earlier cohorts are significantly smaller than in the baseline economy. For instance, the cohort that is most negatively affected by the FF reform suffers a loss of 3.8% in the low wage growth economy, compared to a 11% loss in the baseline economy. The low-discount planner would now prefer the FF reform over any of the alternatives – the welfare gain arising from the reduction in the tax wedge. Finally, the large welfare gains from the PAYGO alternative reform by and large vanish. The low-discount planner would now prefer the benchmark reform to the PAYGO reform.

2.5.1.2 Scenario 2: Slower convergence

In this scenario, we assume an annual wage growth of 4% until 2049, and 2% thereafter. From 2049 and onward, the wage gap between China and the US is 57%, as in the baseline scenario.

The results of this experiment are quantitatively similar to the baseline case. Delaying the reform until 2051 requires lowering the replacement rate (relative to the benchmark 2013 reform) by 3 percentage point, as in the baseline wage growth scenario. The mid-left panel of figure 2.11 plots the welfare gains/losses of generations retiring between 2000 and 2110 in the case of a delay of the reform until 2050. The continuous line refers to the slow wage growth scenario, whereas the dashed line refers to the baseline wage growth scenario, for comparison. From the social planner's standpoint, the net effect of delaying the reform is about the same: delaying the reform till 2050 delivers a consumption-equivalent welfare

gain to the low (high) discount planner of 1.0% (6.2%), approximately the same as in the baseline scenario.[38]

The distribution of welfare gains in the FF and PAYGO experiments are essentially the same as in the baseline economy (see mid-center and mid-right panels of figure 2.11). The PAYGO reform continues to dominate over all alternative options: a gain of 1.8% (12.6%) accrues the low (high) discount planner, compared to 1.7% (12.8%) in the baseline case.

2.5.1.3 Summary

In summary, wage convergence, a typical feature of emerging economies, is critical for the welfare gains of delaying a reform (or of switching to PAYGO as opposed to an FF system). It is the convergence per se rather than its speed that matters. We have considered two scenarios in which the average Chinese wage converges to 57% of the US level (an assumption that we regard as realistic, if conservative). In one case, convergence ends in 2039, in another it takes until 2069. The welfare implication of the alternative pension reforms considered are essentially identical. In contrast, the results are very different, even reversed, if we shut down the process of wage convergence. The comparison with a constant 2% wage growth scenario is especially revealing, since it is consistent with the standard assumption for pension analyses of developed economies.

2.5.2 Lower fertility

Our forecasts are based on the assumption that the TFR will increase to 1.8 already in 2013. This requires a reform or a softer implementation of the current one-child policy. In this section, we consider an alternative lower fertility scenario along the lines of scenario 1 in Zeng (2007). In this case, the rural and urban TFRs are assumed to be 1.98 and 1.2 forever, implying an ever-shrinking total population. We view this as a lower bound to reasonable

[38] While for simplicity we have reported the welfare gain of delaying the reform until the same year as was optimal in the baseline scenario, with slower convergence the low-discount planner would find it optimal to delay the reform until 2050. The reason is that there are now more poor generations, and thus the planner would want to retain for longer the old generous replacement rate. The result would be similar if one considered this alternative approach.

fertility forecasts. Next, we consider the welfare effects of the two alternative reforms. The three bottom panels of figure 2.11 show the welfare effects.

Under this low-fertility scenario, the benchmark reform requires an even more draconian adjustment. The replacement rate must be set equal to 35.5% as of 2013. Delaying the reform is now substantially more costly. A reform in 2050 requires a replacement rate of 23.7%. The trade-off between current and future generations becomes sharper than in the baseline economy. On the one hand, there are larger gains for the cohorts retiring between 2013 and 2050 relative to the benchmark reform. On the other hand, the delay is more costly for the future generations. Aggregating gains and losses yields a gain for the low-discount planner of 3.0%, significantly larger than in the baseline economy. The FF reform exhibits larger losses than in the baseline model (even the low-discount planner prefers the benchmark to a fully funded reform). Moreover, the PAYGO reform yields larger gains than in the benchmark reform.

The reason for the larger gains from delaying the reform or switching to PAYGO is related to the fact an economy with a low population growth is intrinsically poorer – which is reflected in a lower replacement rate in the benchmark case. Thus, sticking to the current rule (60% replacement rate for the earlier generations) implies more intergenerational redistribution than in the baseline economy. Since we have shown that the planner would like substantially replacement rates than 60% for the transition generation (see figure 2.6), the low population growth relaxes the constraint for the planner. Thus the planner has a stronger preference for the delayed reform than in the richer baseline economy. A similar argument apply to the PAYGO reform.

2.5.3 Slower migration

In the baseline case, the future age-specific migration rates are assumed to be time invariant. One might find it plausible that as urbanization proceeds, the migration rates will dwindle. We considered the following alternative experiment: we scaled down all migration rates to 55.5% of the baseline rates. This implies that the urban share of the total population is 67.7% in 2050, compared to 80.8% in the baseline economy. We view this as a lower bound to a realistic description of the migration process. The results are quantitatively similar. In

fact, the adjustment of the replacement rate required to achieve financial sustainability is slightly lower under slow migration than in the baseline scenario (reduction of replacement rate to 40.2% vs. 40.0%). Intuitively, in the initial years (i.e., until 2038) the migration flow is larger in the baseline scenario. However, after 2039 the slow migration scenario implies a larger migration flow (i.e., migrants per year), since more people are left in the rural areas. Thus, in the slow migration scenario more migrants enter the urban sector when wages are already high and wages and pensions grow slowly. This makes for a larger contribution to the pension system than does a massive migration in the first period, when productivity is still low and wage and pension growth are higher.

The comparison between alternative reforms yields similar results to the baseline model (see Table 1).

2.5.4 High interest rate

In the macroeconomic literature on pension reforms in developed economies, it is common to assume that the return on the assets owned by the pension fund is equal to the marginal return to capital. In this paper, we have calibrated the return on assets to 2.5%. However, the empirical rate of return on capital in China has been argued to be much higher (see discussion above). To get a sense of the role of this assumption, we now consider a scenario in which the interest rate is much higher – equal to 6% – between 2013 and 2050.

There are two main differences between the scenarios with lower and higher interest rates. First, delaying the reform yields much smaller gains for the transitional generations, and in fact the low-discount planner is essentially indifferent between the benchmark reform and a delay until 2051. Second, the FF reform entails larger gains for the future generations and smaller losses for the current generations relative to the baseline calibration. As should be expected, when the interest rate is significantly higher than the average growth rate, the PAYGO system becomes less appealing, because the gains to current generations are smaller. In particular, the low-discount planner prefers the FF to the PAYGO reform.

2.5.5 Perfect foresight about alternative reforms

In the main analysis, each alternative policy reform came as a "surprise" to agents in 2013. The main insights are unchanged if one assumes, alternatively, that all reforms are perfectly anticipated in year 2000. The main difference is that the welfare gains of all alternative reforms are slightly larger, because households tailor their saving behavior prior to 2013 to the realized policy outcome. See table 2.1.

2.6 A dynamic general equilibrium model

Up to now, we have taken the wages and the rate of return on savings as exogenous. The normative predictions of the previous sections hinge on the assumed wage path. The key feature is a transition implying a temporarily high wage growth, characteristic of emerging economies. The wage path was assumed to be exogenous. In this section, we construct a dynamic general equilibrium model that delivers the wage and interest rate sequence assumed in the baseline model of section 2.3 as an equilibrium outcome. These prices are sufficient to compute the optimal decisions of workers and retirees (consumption and labor supply) as well as the sequence of budget constraints faced by the government. The model is closely related to Song et al. (2011), augmented with the demographic model of section 2.3.2 and the pension system of section 2.3.

The production sector: The urban production sector consists of two types of firms: (i) *financially integrated* (F) firms, modeled as standard neoclassical firms; and (ii) *entrepreneurial* (E) firms, owned by (old) entrepreneurs, who are residual claimants on the profits. Entrepreneurs delegate the management of their firms to specialized agents called *managers*. E firms can run more productive technologies than F firms (see Song et al., 2011 for the microfoundation of this assumption). However, they are subject to credit constraints that limit their growth. In contrast, the less productive F firms are unconstrained. Motivated by the empirical evidence (see Song et al., 2011) that private firms are more productive and more heavily financially constrained than state-owned enterprises (SOE) in China, we think of F firms as SOE and E firms as privately owned firms.

The technology of F and E firms are described, respectively, by the following production functions:

$$Y_F = K_F^{\alpha} \left(AN_F\right)^{1-\alpha}, \qquad Y_E = K_E^{\alpha} \left(\chi AN_E\right)^{1-\alpha},$$

where Y is output and K and N denote capital and labor, respectively. The parameter $\chi > 1$ captures the assumption that E firms are more productive. A labor market-clearing condition requires that $N_{E,t} + N_{F,t} = N_t$, where N_t denotes the total urban labor supply at t, whose dynamics are consistent with the demographic model. The technology parameter A grows at the exogenous rate z_t; $A_{t+1} = (1+z_t)A_t$.

The capital stock of F firms, $K_{F,t}$, is *not* a state variable, since F firms have access to frictionless credit markets, and the capital stock adjusts so that the rate of return on capital equals the lending rate. Note that we assume no irreversibility in investments, so F firms can adjust the desired level of capital in every period. Let r_t^l denote the net interest rate at which F firms can raise external funds. Let w denote the market wage. Profit maximization implies that $K_F = AN_F \left(\alpha / \left(r_t^l + \delta\right)\right)^{-\frac{1}{1-\alpha}}$, where δ is the depreciation rate. The capital-labor ratio and the equilibrium are determined by r^l. Thus,

$$w_t \geq (1-\alpha) \left(\frac{\alpha}{r_t^l + \delta}\right)^{\frac{\alpha}{1-\alpha}} A_t. \tag{2.20}$$

As long as there are active F firms in equilibrium ($N_F > 0$), equation (2.20) holds with strict equality.

Let $K_{E,t}$ denote the capital stock of E firms. E firms are subject to an agency problem in the delegation of control to managers. The optimal contract between managers and entrepreneurs requires revenue sharing. We denote by ψ the share of the revenue accruing to managers.[39] Profit maximization yields, then, the following optimal labor hiring decision:

$$\begin{aligned} N_{Et} &= \arg\max_{\tilde{N}_t} \left\{ (1-\psi) \left(K_{Et}\right)^{\alpha} \left(\chi A_t \tilde{N}_t\right)^{1-\alpha} - w_t \tilde{N}_t \right\} \\ &= \left((1-\psi)\chi\right)^{\frac{1}{\alpha}} \left(\frac{r_t^l + \delta}{\alpha}\right)^{\frac{1}{1-\alpha}} \frac{K_{Et}}{\chi A_t}. \end{aligned} \tag{2.21}$$

[39] Managers have special skills that are in scarce supply. If a manager were paid less than a share ψ of production, she could "steal" it. No punishment is credible, since the deviating manager could leave the firm and be hired by another entrepreneur. See Song et al. (2011) for a more detailed discussion.

The gross rate of return to capital in E firms is given by

$$\rho_{E,t} = \left((1-\psi) K_{Et}^{\alpha} (\chi A_t N_{Et})^{1-\alpha} - w_t N_{Et} + (1-\delta) K_{Et} \right) / K_{E,t}. \tag{2.22}$$

We assume that E firms are also subject to a credit constraint, modeled as in Song et al. (2011, p. 216). According to such a model, E firms can borrow funds at the same interest rate as F firms, but the incentive-compatibility constraint of entrepreneurs implies that the share of investments financed externally must satisfy the following constraint:

$$K_E - \Omega_{E,t} \leq \frac{\sigma \rho_E}{1+r^l} K_E, \tag{2.23}$$

where $\Omega_{E,t}$ denotes the stock of entrepreneurial wealth invested in E firms at t, and, hence, $K_E - \Omega_{E,t}$ denotes the external capital of E firms. Thus, the constraint implies that the entrepreneurs can only pledge to repay a share σ of next-period net profits.

Three regimes are possible: (i) during the first stage of the transition, the credit constraint (2.23) is binding and F firms are active (hence, the wage is pinned down by (2.20) holding with equality); (ii) during the mature stage of the transition, the credit constraint (2.23) is binding and F firms are inactive; (iii) eventually, the credit constraint (2.23) ceases to bind (F firms remain inactive). In regimes (ii) and (iii), (2.20) holds with strict inequality.

Consider, first, regime (i). Substituting N_{Et} and w_t into (2.22) by their equilibrium expressions, (2.20) and (2.21), yields the gross rate of return to E firms: $\rho_{E,t} = (1-\psi)\left((1-\psi)\chi\right)^{\frac{1-\alpha}{\alpha}} \left(r_t^l + \delta\right) + (1-\delta)$. The corresponding gross rate of return to entrepreneurial investment is given by $R_{E,t} = \left(\rho_{E,t} K_{E,t} - \left(1+r_t^l\right)\left(K_{E,t} - \Omega_{E,t}\right)\right) / \Omega_{E,t}$. We assume that $(1-\psi)^{\frac{1}{\alpha}} \chi^{\frac{1-\alpha}{\alpha}} > 1$, ensuring that the return to capital is higher in E firms than in F firms (i.e., that $R_{E,t} > r_t^l + 1$). Note that the rate of return to capital is a linear function of r_t^l in both E and F firms. The equilibrium in regime (i) is closed by the condition that employment in the F sector is determined residually, namely,

$$N_{F,t} = N_t - \left((1-\psi)\chi\right)^{\frac{1}{\alpha}} \left(\frac{r_t^l + \delta}{\alpha} \right)^{\frac{1}{1-\alpha}} \frac{K_{Et}}{\chi A_t} \geq 0.$$

Consider, next, regime (ii), where only E firms are active ($N_{E,t} = N_t$) and the borrowing constraint is binding, so (2.23) holds with equality. In this case, the rates of

return to capital and labor equal their respective marginal products. More formally, $w_t = (1-\alpha)(1-\psi)(\chi A_t)^{1-\alpha}(K_{E,t}/N_t)^{\alpha}$, and the gross rate of return on entrepreneurial wealth is given by

$$\rho_{E,t} = \left(\alpha(1-\psi)\chi^{1-\alpha}\left(\frac{K_{Et}}{A_t N_t}\right)^{\alpha-1} + (1-\delta)\right),$$

whereas the borrowing constraint implies that $K_{E,t} = \left(1 + \frac{\sigma\rho_{E,t}}{R^l - \sigma\rho_{E,t}}\right)\Omega_{E,t}$. Given the stock of entrepreneurial wealth, $\Omega_{E,t}$, the two last equations pin down $\rho_{E,t}$ and $K_{E,t}$. The rate of return to entrepreneurial investment is then determined by the expression used for regime (i).

Finally, in regime (iii) the rate of return to capital in E firms is identical to the rate of return offered by alternative investment opportunities (e.g., bonds). Namely,

$$R_{E,t} = 1 + r_t^l.$$

Thus, $K_{E,t}$ ceases to be a state variable, and the wage is given by $w_t = (1-\alpha)\left(\alpha/\left(r_t^l + \delta\right)\right)^{\alpha/(1-\alpha)}\chi A_t$.

In all regimes, the law of motion of entrepreneurial wealth is determined by the optimal saving decisions of managers and entrepreneurs, described below.

The rural production sector consists of rural firms whose technology is assumed to be similar to that of urban F firms, $Y_{Rt} = K_{Rt}^{\alpha_R}(\chi_R A_t N_{Rt})^{1-\alpha_R}$, where $\chi_R < 1$. Like urban F firms, rural firms can raise external funds at the interest rate r_t^l in each period, and adjust their capital accordingly. So, r_t^l pins down capital-labor ratio and wage in the rural economy. This description is aimed to capture, in a simple way, the notion that there are constant returns to labor in rural areas, due to, e.g., rural overpopulation.

Banks: Competitive financial intermediaries (*banks*) with access to perfect international financial markets collect savings from workers and hold assets in the form of loans to domestic firms and foreign bonds. Foreign bonds yield an exogenous net rate of return denoted by r, constant over time. Arbitrage implies that the rate of return on domestic loans, r_t^l, equals the rate of return on foreign bonds, which in turn must equal the deposit rate. However, lending to domestic firms is subject to an *iceberg cost,* ξ, which

captures the operational costs, red tape, and so on, associated with granting loans. Thus, ξ is an inverse measure of the efficiency of intermediation. In equilibrium, $r^d = r$ and $r_t^l = (r + \xi_t) / (1 - \xi_t)$, where r_t^l is the lending rate to domestic firms.

Households' saving decisions: Workers and retirees face the problem discussed in section 2.3, given the equilibrium wage sequence, and having defined $R \equiv 1 + r$. As in the previous section, we hold fixed the share of workers participating in the pension system.

The young managers of E firms earn a managerial compensation m. Throughout their experience as managers, they acquire skills enabling them to become entrepreneurs at a later stage of their lives. The total managerial compensation in period t equals $M_t = \psi Y_{E,t}$. Managers work for J_E years, and during this time can only invest their savings in bank deposits (as can workers) which yields an annual gross return R. As they reach age $J_E + 1$, they retire as managers, and have option (which they always exercise) to become entrepreneurs. In this case, they invest their wealth in their own business yielding the annual return $R_{E,t}$ and hire managers and workers. Thereafter, they are the residual claimants of the firm's profits. We assume that entrepreneurs are not in the pension system. Their lifetime budget constraint is then given by

$$\sum_{j=0}^{J_E} \frac{s_j}{R^j} c_{t+j} + \sum_{j=J_E+1}^{J} \frac{1}{R^{J_E}} \frac{s_j}{\Pi_{v=t+J_E+1}^{t+j} R_{E,v}} c_{t+j} = \sum_{j=0}^{J_E} \frac{s_j}{R^j} m_{t+j}.$$

The right hand-side is the PDV income from the managerial compensation. The left hand-side yields the PDV of consumption. This is broken down in two parts: the first term is the PDV of consumption when young, when the manager faces a constant rate of return, R; the second part is the PDV of consumption when being an entrepreneur, and is discounted at the rate R till J_E, and at the entrepreneurial rate of return thereafter.

Mechanics of the model: The dynamic model is defined up to a set of initial conditions including the wealth distribution of entrepreneurs and managers, the wealth of the pension system, the aggregate productivity (A_0), and the population distribution. The engine of growth is the savings of managers and entrepreneurs. If the economy starts in regime (i), then all managerial savings are invested in the entrepreneurial business as soon

as each manager becomes an entrepreneur. As long as managerial investments are sufficiently large, the employment share of E firms grows and that of F firms declines over time.

The comparative dynamics of the main parameters is as follows: (i) a high β implies a high propensity to save for managers and entrepreneurs and a high speed of transition; (ii) a high world interest rate (r) and/or a high iceberg intermediation cost (ξ) increases the lending rate, implying a low wage, a high rate of return in E firms, a high managerial compensation, and, hence, a high speed of transition; (iii) a high productivity differential (χ) implies a high rate of return in E firms, a high managerial compensation, and, hence, a high speed of transition; (iv) a high σ implies that entrepreneurs can leverage up their wealth and earn a higher return on their savings, which speeds up the transition; and (v) a high managerial rent (ψ) implies a low rate of return in E firms, a high managerial compensation, and, hence, has ambiguous (and generally non-monotonic) effects on the speed of transition.

Note that the savings of the worker do not matter for the speed of transition, because the lending rate offered by banks depends only on the world market interest rate and on the iceberg cost.

2.6.1 Calibration

In Song et al. (2011), we show that a calibrated version of the model outlined in the previous section matches well a number of salient macroeconomic trends for the recent period. In particular, the model reproduces realistic trends for output growth, wage growth, return to capital, transition from state-owned to private firms, and foreign surplus accumulation. The current model - which incorporates additional features including demographics and the pension system - the model is calibrated to match the same macroeconomic trends after 2000.

We must calibrate two parameters related to the financial system, ξ and σ, and four technology parameters, α, δ, χ and ψ. The parameters α and δ are set exogenously: $\alpha = 0.5$ so that the capital share of output is 0.5 in year 2000 (Bai et al., 2006), and $\delta = 0.1$ so that the annual depreciation rate of capital is 10%.

The remaining parameters are calibrated internally, so as to match a set of empirical moments. We set the parameters ψ and χ so that the model is consistent with two key observations: (i) the capital-output ratio in E firms is 50% of the corresponding ratio in F firms (as documented by Song et al. (2011) for manufacturing industries, after controlling for three-digit industry type), (ii) the rate of return on capital is 9% larger in E firms than in F firms.[40] The implied parameter values are $\psi = 0.27$ and $\chi = 2.73$. This implies that the TFP of an E firm is 1.65 times larger than the TFP of an F firm.[41]

We set ξ so as to target an average gross return on capital of 20% in year 2000 (Bai et al., 2006). With $\delta = 10\%$, this implies an average net rate of return on capital of 10%. This average comprises both F firms and E firms. Since the DPE employment share in the period 1998-2000 was on average 10%, this implies $\rho_F = 9.3\%$, so that the initial value for ξ is $\xi_{2000} = 0.062$. After year 2000, we assume that there is gradual financial improvement so ξ falls linearly to zero by year 2024. The motivation for such decline is twofold. First, we believe it is reasonable that banks improve their lending practices over time, so that borrowing-lending spreads will eventually be in line with corresponding spreads in developed economies. Second, a falling ξ will generate capital deepening in F firms and E firms due to cheaper borrowing and higher wages, respectively. Such development helps the model to generate an increasing aggregate investment rate during 2000-2009, which is a clear pattern of aggregate data. If ξ were constant, the model would predict a falling rate (see Song et al., 2011, for further discussion).

We set $\sigma = 0.43$, so that entrepreneurs can borrow 87 cents for each dollar in equity in 2000. This value for σ implies that the growth in the DPE employment share is in line with private employment growth between 2000 and 2008 in urban areas. We set the initial level of productivity, A_{2000}, so that the urban GDP per capita is 20% of the US level in 2011. Moreover, we set the growth rate of A_t (i.e., the secular exogenous productivity growth) so that the model generates an aggregate growth in GDP per capita of 9.7% for China during 2000-2011. The resulting growth rate in A_t is 2% larger than the associated world growth

[40] Song et al. (2011) document that manufacturing, domestic private enterprises (DPE) have on average a ratio of profits per unit of book-value capital 9% larger than that of SOEs during the period 1998-2007. A similar difference in rate of return on capital is reported by Islam, Dai, and Sakamoto (2006).

[41] Hsieh and Klenow (2009) estimate TFP across manufacturing firms in China and find that the TFP of DPEs is about 1.65 times larger than the TFP of SOEs.

rate during this period. After 2011, this excess growth in A_t falls linearly to zero until the TFP level in E firms is equal to that of US firms. This occurs in year 2022. Finally, β is calibrated to 1.0179 to match the average aggregate saving rate of 48.2% in 2000-2010.

In the rural sector, we set $\alpha_R = 0.3$ to match the observed 20% investment rate in the rural area in 2000. The technology gap χ_R is set to 0.75 to capture an observed urban-rural wage gap of 1.84 in 2000. The rural wage grows over time, due to the exogenous technology growth and to the decreasing lending rate. The rural-urban wage gap implied by the model increases from 1.84 in 2000 to 3.47 in 2040 and stays constant thereafter.

The initial conditions are set as follows. Total entrepreneurial wealth in 2000 is set equivalent to 14.6% of urban GDP so that the 2000 DPE employment is 20%. The distribution of that entrepreneurial wealth is obtained by assuming that all entrepreneurs are endowed with the same initial wealth in 1992 (1992 is the year when free-market reforms in China accelerated). Moreover, all managers are assumed to start with zero wealth in 1992. Initial wealth for workers and retirees is also set to zero in 1992. The 2000 distribution of wealth across individuals is then derived endogenously. Finally, the initial government wealth is set to 71% of GDP in 2000 so as to generate a net foreign surplus equal to 12% of GDP in 2000.[42]

2.6.2 Simulated output trajectories

The calibrated model yields growth forecasts that we view as plausible. Figure 2.12 shows the evolution of productivity and output per capita forecasted by our model. The growth rate of GDP per worker remains about 8.5% per year until 2020 (see upper panel). After 2020, productivity growth is forecasted to slow down. This is driven by two forces: (i) the end of the transition from state-owned to private firms and (ii) the slowdown in technological convergence. The growth rate remains above 6.9% between 2020-2030 and eventually dies off in the following decade. Note that the growth of GDP per capita is lower than that of GDP per worker after 2015, due to the increase in the dependency ratio. On average, China is expected to grow at a rate of 6.5% between 2013 and 2040. The contribution of human

[42] More precisely, government wealth is calculated as a residual. It is equal to the sum of foreign surplus and domestic capital (from both SOE and DPE) minus the stock of private wealth owned by workers and entrepreneurs.

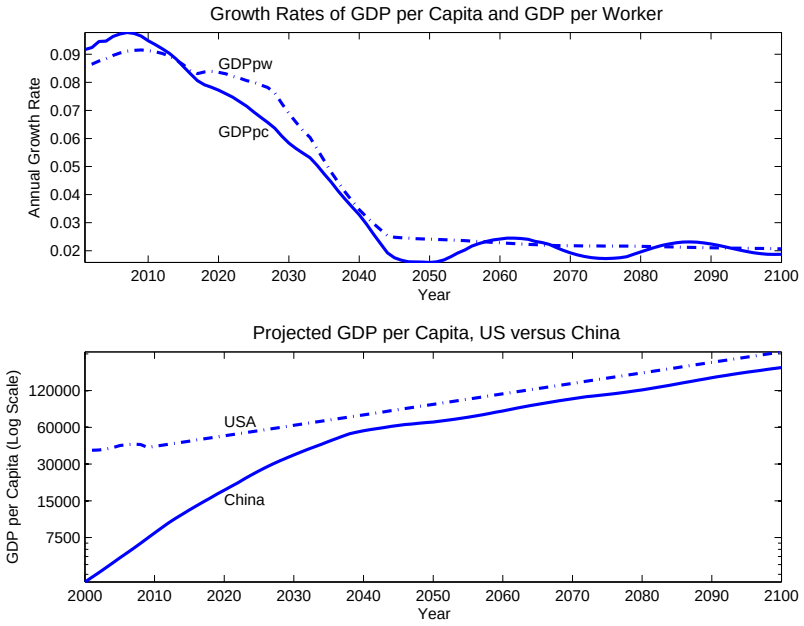

Figure 2.12: GDP

The upper panel shows projected annual growth rates in GDP per worker and GDP per capita in the calibrated economy. The lower panel shows projected GDP per capita in levels for China and the US.

capital is 0.8% per year, due to the entry of more educated young cohorts in the labor force. In this scenario, the GDP per worker in China will be 73% of the US level by 2039, remaining broadly stable thereafter. Total GDP in China is set to surpass that in the United States in 2013 and to become more than twice as large in the long run.

The wage sequence that was assumed in section 2.3 is now an endogenous outcome. Wages are forecasted to grow at an average of 5.1% until 2030 and to slow down thereafter. What keeps wage growth high after 2020 is mostly capital deepening.[43]

[43] In Section 2.4 we held the wage sequence constant across the different policy experiments. However, in the general equilibrium model of this section, the wage sequence is endogenous and would in general be affected by alternative reforms. In particular, pension reforms impact labor supply through a wealth effect, and this influences the capital accumulation dynamics during transition. Since the effects are quantitatively small, the results are omitted and are available upon request.

2.6.2.1 Sensitivity: high savings and foreign surplus

Although the growth forecasts are plausible, the calibrated economy generates a very large amount of savings. For instance, in 2070 the economy has a wealth-GDP ratio exceeding 1000%. This is because the model is calibrated to match aggregate savings during 2000-2010. In that period, China experienced high growth and yet a very high saving rate (48.2% on average).

Since our stylized model forecasts an eventual decline in growth, the intertemporal motive would suggest that consumption should have been high before 2010. Therefore, the model requires a sufficiently high discount factor ($\beta = 1.0179$) in order to predict the empirical saving rate during the first decade of the 21st century. In our model, a high β is a stand-in for a number of institutional features that are not explicitly considered and that may explain a high propensity to save over and beyond pure preferences (e.g., large precautionary motives or large downpayment requirements for house purchases).[44]

Since it seems implausible that China will continue to save so much, we consider an alternative scenario, where all cohorts entering the labor market after 2013 have $\beta = 0.97$. In such an alternative scenario China's net foreign position would be zero in the long run. The analysis of the alternative pension arrangements discussed in the previous sections yields essentially the same results as in the high β economy. Thus, the calibration of β is unimportant for the effects of the welfare analysis, which is the main contribution of this paper.

This finding is not surprising since long-term wages and GDP do not hinge on the domestic propensity to save. Although the entrepreneurs' propensity to save determines the speed of the transition, this does not to matter much for welfare (see section 2.5.1).

2.6.2.2 Sensitivity: Financial development

The model borrows from Song et al. (2011) the assumption that E firms are financially constrained. Note that the salience of the financial constraints declines over time as E firms

[44] Chamon et al. (2010) and Song and Yang (2010) study household savings in calibrated life-cycle models. They incorporate individual risk and detailed institutional features of the pension system and find that their models are qualitatively consistent with the life-cycle profile of household saving rates. However, both studies find that with a conventional choice of β, their models would imply quantitatively too low savings for the young households.

accumulate capital. As the economy enters regime (iii), which occurs in 2038, the financial constraint ceases to bind.

In our baseline calibration, the parameter σ, which regulates borrowing of private firms, is assumed to be constant over time. An exogenous increase in σ – for example, due to financial development – would speed up growth of private firms. Wage growth would accelerate earlier, although the long-run wage level would be unaffected.

To study the effects of financial development on pension reform, we consider a stark experiment in which the borrowing constraint on private firms is completely removed in 2013. This means that state-owned firms vanish, and there is large capital inflow driven by entrepreneurial borrowing. Wages jump upon impact (by 85%) due to the large capital deepening. In 2030, the wage level is still 15.8% above the baseline calibration. In 2038 the wage level is the same as in the benchmark calibration.

Although financial development affects the transition path, it brings little change to the conclusions of the welfare analysis.[45] The benchmark reform requires a slightly smaller reduction of the replacement rate: 40.7% instead of 40%. The delayed reform still entails gains for the transition cohorts, albeit these gains decline faster over time. For instance, delaying a reform until 2050 yields a 16.5% consumption equivalent gain for the cohort retiring in 2013, but only a 9.7% gain for the cohort retiring in 2051. The losses suffered by the cohorts retiring after 2051 are comparable in size to those in the baseline scenario without financial development. The gains accruing to the high- and low-discount planners are, respectively, 5.3% and 0.5% (6.4% and 0.9% in the baseline scenario).

The FF reform yields slightly better outcomes. All generations retiring after 2050 gain from the reform (2059 in the baseline scenario), and the losses of the earlier cohorts only reach 7% (11% in the baseline scenario). The high-discount planner continues to prefer the benchmark reform to the FF reform, whereas the low-discount planner continues to have the opposite ranking. The PAYGO reform yields even larger gains to the earlier cohorts. Both the high- and the low-discount social planners continue to prefer the PAYGO reform to any alternative policy-driven reform. However, the welfare gap between the PAYGO and the fully funded reform is now smaller, since the planners dislike the concentrated nature

[45] We focus for simplicity on the policy-driven reforms, and we omit an explicit analysis of the optimal policy.

of the gains under the PAYGO reform. For instance, the consumption equivalent gain of the low-discount planner relative to the benchmark reform is 1%, compared with 1.8% in the baseline scenario. Since the fully funded reform also entails a 0.6% gain relative to the benchmark reform, the consumption equivalent gain of the PAYGO relative to the FF reform is only 0.4% (although it remains significantly higher, 12.7%, for the high-discount planner).

In conclusion, financial development mitigates but does not change the welfare implications of alternative reforms.

2.7 Conclusions

Pension systems have been a key instrument for sharing high growth across generations in Western economies after World War II and could potentially play the same role in emerging countries. However, the prospect of an adverse demographic transition threatens the fiscal sustainability of non-funded pension systems. In this paper, we have analyzed the positive and normative effects of alternative pension reforms with the aid of a dynamic general equilibrium model calibrated to China.

A number of studies before us have argued that China must reform its pension system to achieve long-run balance (see, e.g., Sin (2005), Dunaway and Arora (2007), Salditt et al. (2007), and Lu (2011)). Our analysis concurs with this view, but shows that rushing into a draconian reform would have large unequalizing effects: it would harm current generations and only mildly benefit future generations. In a fast-growing society like China, this would imply dispensing with a powerful institution redistributing resources from richer future generations to poorer current generations. Even a planner with an annual discount rate as low as 0.5% would prefer an unfunded pay-as-you-go system to both an immediate sustainable reform and to a reform that pre-funds the pension system. These results obtain in a standard OLG model that predict that, in a mature economy with steady wage growth and perfect capital markets, a fully funded system outperforms an unfunded PAYGO system. These contrasting results highlight the general principle (see, Acemoglu, Aghion and Zilibotti, 2006) that mechanically transposing policy advice from mature to developing or emerging economies may be misleading.

Our study has abstracted from a number of features that can influence the welfare effects of alternative reforms. First, we considered neither idiosyncratic nor intergenerational risk. Both sources of risk are important and difficult to insure in emerging economies, strengthening the case for a non-funded pension system (see, e.g., Krueger and Kuebler 2006, and Nishiyama and Smetters 2007). Second, we abstracted from within-cohort inequality. In reality, public pensions provide also some intragenerational redistribution. Last but not least important, we have abstracted from altruism within families. Public pensions could crowd out private transfers from children to elderly, reducing the social value of pensions. Cai, Giles and Meng (2006) document that retirees in urban China do receive transfers from their children, and that these respond to negative income shocks (e.g., pension arrears). However, such transfers appear to provide only very limited insurance. For instance, when income is close the poverty line, a one yuan temporary reduction in income leads to an increase in net transfers between 10 to 16 cents. Their study concludes that improving the public pension system is unlikely to lead to any significant crowd out of private transfers. This conclusion is shared by Park *et al.* (2012) who add that the effectiveness of the informal private insurance system is set to decline in future (as it did, for instance, in the recent history of Latin America), since elderly will have fewer children and more of them will live separately from their children (see also Yang and Chen, 2010, and Calvo and Williamson, 2008).

Chapter 3

Divide and Rule: An Origin of Polarization and Ethnic Conflict[1]

3.1 Introduction

Violent conflict causes enormous costs for people, yet we often observe wide support in the population for political elites that initiate conflicts. A recent literature has argued theoretically and empirically that ethnic polarization is an important determinant of the incidence of conflict (Esteban, Mayoral, and Ray, 2012). Many recent civil wars were indeed fought along ethnic lines. Conflict between Hutu and Tutsi in Rwanda have led to around 800,000 deaths from genocide and triggered further violence in the form of counter-attacks and civil wars in the larger region (Prunier, 2009).[2] For the vast majority of people the consequences of this ethnic violence were catastrophic.

How can ethnic groups be so highly polarized that deadly violence erupts on a large scale? Predetermined differences can hardly be an explanation, as the same groups have lived in peace for long periods of time. For example, there is no evidence of any political violence between Hutu and Tutsi before 1959 (Gourevitch, 2011) and the ethnic labels did

[1]We would like to thank Dominic Rohner, Viktor Tsyrennikov, Fabrizio Zilibotti, and seminar participants at the University of Zurich for valuable comments. Caroline Chouard and Sebastian Ottinger provided excellent research assistance.

[2]Ethnicity is also a key component of conflicts in Yugoslavia. Section 3.3 provides a more detailed description of these cases.

not even have a political meaning until the mid 19th century (Pottier, 2002). In contrast, there are numerous examples in Rwanda and elsewhere showing that violence was a *strategic choice* of a political elite. Fearon and Laitin (2000) emphasize this in their review of conflict studies:

> "If there is a dominant or most common narrative in the texts under review, it is that large-scale ethnic violence is provoked by elites seeking to gain, maintain, or increase their hold on political power."

By initiating conflict along ethnic lines, elites can deepen ethnic divisions and thus increase polarization (Horowitz, 2000). division in society strengthens the position of the elite and allows it to exploit power. This strategy is well known as "divide-and-rule".

We propose a model of divide-and-rule where the polarization between groups is endogenous. Polarization is modeled as a lack of trust between ethnic groups. High levels of trust threaten the autocratic elite because it may be overthrown if trust is high and the common interest among the ethnic groups in society becomes large. A key question for such a divide-and-rule argument is how an elite can induce people to polarize. Fearon and Laitin (2000) describe this challenge as follows:

> "The puzzle for such theoretical arguments is to explain how elites can convince their followers to adopt false beliefs and take actions that the followers would not want to take if they understood what the leaders were up to."

Our contribution is to provide a micro-foundation for how elites can polarize society when people are rational. This is achieved by embedding a model of trade and trust (Rohner, Thoenig, and Zilibotti, 2013a) into a political economy framework that allows for war and revolution (Besley and Persson, 2011).

The model society has two ethnic groups of equal size, one of which initially being the incumbent group. Within each group, there is a small political elite that derives rents from being in power. We start from a situation of autocracy where the elite of the incumbent group sets policies. The incumbent elite faces a threat from the other group (transition of power) and from the people of its own group (revolution). A revolution leads to a regime switch to democracy where people of both groups trade with each other without being

expropriated by an elite. The threat of revolution therefore depends on the expected gains from trade.

Similar to Rohner, Thoenig, and Zilibotti (2013a), we model the expected outcome of trade as a function of trust between trade partners. More specifically, the outcome of trade is stochastic and the likelihood of a good trade outcome depends on how mutually beneficial trade is (e.g. the complementarity between the trade partners), which is unknown. Agents hold a prior belief about how beneficial trade is. A belief that trade is beneficial implies a high level of trust. Trust can increase through Bayesian updating of this belief after good trade outcomes occur. However, trade can only take place during peaceful times such that trust cannot emerge while there is war. The ruling elite can therefore prevent trust from emerging by starting a war. This enables the elite to limit the threat of revolution that originates in the common interest of people when trust is high.

The precise tradeoffs faced by the people and by the elite are as follows. A currently ruling elite taxes people in both groups but redistributes part of the revenues as transfers to the people of its own group. When trust is very low and therefore expected gains from trade small, then transfers are relatively more important for the people of the incumbent group than the gains from trade. They therefore prefer their own elite to stay in power. The probability that the currently ruling group stays in power is higher when it fights a war against the other group than when it is in peace. Elite and people of the incumbent group therefore both support war when trust is very low. When trust increases somewhat, the people start to prefer peace because they want to reap the gains from trade with the other group. But the elite still prefers war over peace because it is afraid of losing power to the other group. The people are less concerned about such a transition of power because they know that they will still have part of the trade surplus even if the other group takes over. This setting therefore generates that the elite stays at war for longer than is optimal for the people, because the elite has more at stake from being in power.

At intermediate levels of trust, both elite and people of the incumbent group support peace and trade with the other group. Both also prefer to maintain the autocracy because it allows them to exploit the other group. The interest of elite and people of the incumbent group are therefore aligned again for these intermediate levels of trust. But when trust increases further to a high level, the interests start to diverge once more: income from

trade surplus becomes so high that it is more important for the people of the incumbent group than the transfers through their elite. The people may therefore start a revolution and establish a peaceful democracy with the other group in order to reap the full benefits from trade without being taxed by an elite. The elite on the other hand loses all its rents when there is a revolution. At high levels of trust, we therefore again get the result that the elite starts a war in order to lower the chances of losing power, while people would prefer to trade in peace.

In summary, an autocratic elite that knows that trust is relatively high has to decide if it should allow for peace or start a war. It faces the following tradeoff: (1) During peace, people can trade across groups and this may (depending on whether trade outcomes are good) generate trade surpluses that the elite can tax. However, good trade outcomes also imply that people update their beliefs such that trust can emerge. The expectation of high gains from trade (based on high trust) increases the threat of revolution, since people may become willing to incur the cost of revolution in order to not have the trade surplus taxed by the elite. (2) If the elite starts a war, then there is no trade and the elite therefore cannot tax the trade surplus. But it also prevents trust from emerging between the two ethnic groups, which limits the risk of a revolution. A key implication of this setup is that the elite can apply a divide-and-rule strategy: when the threat of revolution is high, it starts a conflict between the groups, which harms trust, polarizes society, and limits the common interest of people. This strategy is against the interest of both groups of people in society, since they may be better off under a democracy where they trade and are not exploited by an elite.

The rest of the paper is organized as follows. Section 3.2 reviews the related literature, including empirical studies on conflict and ethnic polarization as well as theoretical work on the divide-and-rule strategy. In section 3.3, we discuss the anecdotal evidence on the divide-and-rule strategy, in particular the cases of Rwanda and Yugoslavia. Section 3.4 presents a simple benchmark model to illustrate how trust emerges through trade and how it affects the interest and behavior of citizens and elite. In section 3.5, the benchmark model is extended to a dynamic setting to study how trust evolves over time and how it is manipulated by the elite. Finally, we conclude in section 3.6.

3.2 Related Literature

This paper is related to several strands in the existing literature. First, we relate to a large literature on civil conflict (Blattman and Miguel, 2010, provide a survey). Fearon and Laitin (2003) estimate that there were about 127 major civil wars between 1945 and 1999 with more than 16 million fatalities overall. They argue that conditions favoring insurgency and guerrilla warfare are important to understand the onset of post-Cold War conflicts. Our analysis is particularly related to the strand in the civil war literature focusing on conflict between ethnic groups. Esteban and Ray (2011), Esteban, Mayoral, and Ray (2012), and Rohner (2011) show that the polarization of societies along ethnic lines is associated with high degrees of conflict. Esteban and Ray (2008) argue that coalitions along ethnic lines are more likely to emerge than along classes. Caselli and Coleman (2013) point out that certain ethnic characteristics that are easily observable (such as skin color) allow to distinguish between winners and losers of a conflict and therefore make starting a conflict along these lines more profitable. Besley and Persson (2011) provide a framework where repression and civil war can depend on polarization between groups. We contribute to this literature by endogenizing ethnic polarization. In our framework, a ruling elite can strategically affect polarization between citizens in order to sustain its own power. Such patterns of "divide and rule" have been described also outside of the economic literature.[3] Related papers have argued that a political elite can expropriate its supporting citizens because of their fear that otherwise an even less favorable elite would take over (Padro i Miquel, 2007; Figueiredo and Weingast, 1997)

These empirical findings on the salience of ethnic conflict are complemented by a set of theoretical papers on divide-and-rule strategies. Posner, Spier, and Vermeule (2010) analyze divide-and-rule strategies in two simple game theoretic models, the Prisoner's Dilemma and the Stag Hunt Game, and consider how an outsider (the elite) that is affected

[3] See for example the review of case studies by Fearon and Laitin (2000). Similarly, Horowitz (2000) writes "By appealing to electorates in ethnic terms, by making ethnic demands on government, and by bolstering the influence of ethnically chauvinist elements within each group, parties that begin by merely mirroring ethnic divisions help to deepen and extend them. Hence the oft-heard remark in such states that the politicians have created ethnic conflict."

by the cooperation of the agents can influence outcomes.[4] They state that two conditions are important for such strategies: "(1) A unitary actor bargains with or competes against a set of multiple actors. (2) The unitary actor follows an intentional strategy of exploiting problems of coordination or collective action among the multiple actors." They consider different ways to apply a divide-and-rule strategy: destroying communication channels, payment of bribes, imposition of penalties, generating distrust, limiting interaction, and mixing agents with heterogeneous interests. Our model is best described by the category "generating distrust". However, Posner, Spier, and Vermeule (2010) implement this by the elite being able to directly manipulate agents beliefs about the other agents' private payoffs. In contrast, we provide a setting where the elite cannot directly manipulate believes, but it can limit the learning process between agents. In this sense our setting also relates to the category "limiting interaction", although their use of this strategy refers to limiting the time horizon of repeated interactions, which makes cooperation less likely.

Acemoglu, Robinson, and Verdier (2004) consider a divide-and-rule strategy that falls into the category described by Posner, Spier, and Vermeule (2010) as "payment of bribes". In Acemoglu, Robinson, and Verdier (2004), elites can exploit citizens because they can prevent them from cooperating in a revolution. The kleptocratic elite achieve this by bribing one of two groups in society in order to induce them to reject offers by the other group to cooperate in a revolution. They show that such kleptocratic regimes are more likely to arise if the regime faces fractionalized opponents instead of large players that can solve the coordination problem internally and thus put a constraint on the behaviour of the ruler.

We model polarization as a lack of common interest, which in turn depends on trust between trade partners. We relate to Rohner, Thoenig, and Zilibotti (2013) where trust and cooperation is shaped by Bayesian updating during interactions with the trade partner. Rohner, Thoenig, and Zilibotti (2013) show that even accidental conflicts can be persistent and generate cycles of recurring conflict due to the destruction of trust. Such cycles are also a feature of Acemoglu, Ticchi, and Vindigni (2010), although without linking it to trade. A related literature has analyzed social learning (for example Banerjee, 1992 and Acemoglu

[4]Posner, Spier, and Vermeule (2010) use the term "divide-and-conquer", which is equivalent to "divide-and-conquer". Oxford Dictionary (2014) defines both terms as "The policy of maintaining control over one's subordinates or opponents by encouraging dissent between them, thereby preventing them from uniting in opposition."

and Wolitzky, 2014). Our contribution to this literature is to show that a conflict can be started strategically by the currently ruling elite in order to affect people's beliefs and to sustain its own power.

The channel through which in this framework polarisation can be manipulated by a political elite is trade. We therefore relate to the literature on the relationship between conflict and trade – without the link to trust. The "liberal peace" argument postulates that increased economic interdependence through trade between countries reduces the likelihood of conflict. However, recent contributions find that this relationship is more complex. Barbieri (1996) finds that strong economic interdependence is positively associated with militarised conflict. Martin, Mayer, and Thoenig (2008) show theoretically and empirically that increases in multilateral trade increases the risk of conflict, especially between neighboring states.[5] In our framework, an increase in trade may fail to prevent conflict for a different reasons: a forward looking elite fears that high levels of economic interaction will increase the common interest of citizens and this bears the threat that the elite may be overthrown by the citizens.

3.3 Evidence on Divide And Rule Strategies

Our model provides a micro-foundation for why and how an elite strategically polarizes society in order to sustain its own power. Such patterns have frequently been described in the literature (Fearon and Laitin, 2000). The evidence seems to suggest that provoking conflict is a way of polarizing society. We discuss two cases below that illustrate the role of violence for polarization. These examples will show that ethnic conflict was initiated by elites with the purpose of dividing society, which allowed them to strengthen their role as the ruling elite. We first discuss the case of Rwanda in detail and then show that similar developments have occurred in Yugoslavia.

[5]The liberal peace argument has recently also been challenged by other authors, for example Goldsmith (2013) or Gowa and Hicks (2014).

Rwanda

The history of Rwanda and of its neighbouring countries contains several episodes of large-scale ethnic violence. The conflict between Tutsi and Hutu is often associated with extreme hatred and violence between deeply divided ethnic groups.[6] But the ethnic cleavages were not just due – for example – to genetic differences between people. Neither is there evidence that ethnic groups in Rwanda have always been in violent conflict with each other. In fact, there has been no systematic political violence between Hutu and Tutsi before 1959 (Gourevitch, 1998).

The history of Rwanda is an example of how ethnicity is to some extent "constructed" by elites for their own benefit.[7] We show below that the motivation for creating ethnic tensions often was the elite's aim to sustain their power in response to political or socio-economic changes. Contrary to the popular belief that predetermined ethnic differences triggered conflict, it seems that the conflict after 1990s "had its origin in modern struggles for power and wealth" (Pottier, 2002).

Evidence on how the terms Hutu and Tutsi were initially used prior to colonisation is scant. But it appears that before 1860 there was substantial social mobility and "ethnicity was not a principal organising factor" (Pottier, 2002).[8] From the mid 19th century on, it is known that the Tutsi king Rwabugiri started to polarise the Rwandan people with discriminating rules concerning the ownership of cattle (Newbury, 1988). rich in cattle were regarded as Tutsi while poorer families were labeled Hutu. Therefore, "[. . .] wealth, not race, was the basis of the ethnic distinction between Hutu and Tutsi" (Pottier, 2002).

This policy is an early example in Rwandan history of strategic polarisation by an autocratic ruler who sought to strengthen his power. her study of the impact on the region of Kinyaga, Newbury (1978) states that "social stratification" was among the most important transformation due to the new rule. The king established chiefs from outside of the community in order to collect taxes and these chiefs were typically Tutsi. This led to a strong association of the sharpened ethnic distinction with social status and a notion of inferiority

[6] Hutu (Bantu) are the majority group in Rwanda with 84% of the population. Tutsi (Hamitic) account for 15% and Twa (Pygmy) for 1% (CIA World Factbook, 2014).

[7] See Fearon and Laitin (2000) for a comprehensive discussion of the "construction" of ethnic differences and the implications for violent conflict.

[8] This evidence is based on Newbury's (1978) study of Kinyaga.

of the Hutus (but it did not yet lead to systematic violence). The use of external chiefs also broke up the ties within clans and led to struggles among the groups to be favoured by the rulers. This weakened the citizens and strengthened the rule of the elite (Newbury, 1978). while ethnicity initially played a minor role in Rwanda, it became an increasingly important factor due to policies that polarised society.

With colonization, ethnic identity gained further relevance. The European colonizers used the Tutsi administration in order to control the country and even helped it in expanding the Tutsis' region of influence. The Hutu were under "dual colonialism" by the Tutsi administration and Belgian colonizers (Newbury, 1998). The Belgium support of the Tutsi elite ended only shortly before independence when it started supporting the Hutu majority (Pottier, 2002). Belgian colonizers actively divided society further. In 1933-34, they introduced identity cards that labeled individuals as either "Hutu", "Tutsi", or "Twa" (Gourevitch, 1998). This labelling drastically reduced the mobility between Hutu and Tutsi, which previously was relatively high. Gourevitch (1998) explains how ethnicity gained importance over the last decades as a consequence of the polarising strategy of the elite:

> "The identity cards made it virtually impossible for Hutus to become Tutsis, and permitted the Belgians to perfect the administration of an apartheid system rooted in the myth of Tutsi superiority. [...] Whatever Hutu and Tutsi identity may have stood for in the precolonial state no longer mattered; the Belgians had made "ethnicity" the defining feature of Rwandan existence.

Rwanda therefore had several episodes that serve as examples for how ethnic differences were constructed by an elite for their own benefit. polarisation of Rwanda through dividing people along ethnic lines had a persistent effect on the relationship between Hutu and Tutsi. It repeatedly led to large scale conflict over several decades in Rwanda and its neighbouring countries. ethnic distinction that was initially strengthened by the king and then by colonial powers was later used by the elites of both Hutu and Tutsi in order to gain political power.

According to Prunier (1995), the ethnic identity was a key aspect in the mobilization of Hutu peasants in the genocide that started in 1994. One channel through which elites could affect polarization was trust. Hutu extremists prior to the start of the genocide purposely fostered conflict in order to harm the interaction between moderate Hutu and Tutsi

and reduce trust. Their success with this strategy of limiting the relevance of moderate Hutus strengthened their position within the Hutu group. This allowed them to mobilize the masses for the genocide. Fearon and Laitin (2000) describe this episode as follows:

> "In 1992, two years before the genocide, moderate Hutus gained some control over the tense situation and negotiated a cease-fire with the Rwandan Patriotic Front (RPF, a guerrilla movement that despite seeking a multiethnic constituency, represented Tutsi interests) at Arusha. But Hutu extremists led by the president's wife, Agathe Habyarimana, began taking to the streets against the ensuing peace process. She and her three brothers helped form the "Zero Network" death squads, the institutional precursors of the genocide. After a formal power-sharing deal was signed in January 1993, and the day the International Commission on Human Rights mission left, the extremist Hutus sent their squads to the northwest region where they were strong, and three hundred Tutsis were killed in six days of violence. The in-exile Tutsi-led army then broke the cease-fire and marched across the Ugandan border toward the Rwanda capital, with many of the soldiers defying their own moderate leadership. These wildcats engaged in counterviolence, scaring many Hutus who escaped to Zaire. [...] to assign blame for the failure of the cease-fire with certainty, Hutu moderates increased their estimate that the RPF could not be trusted in political negotiations, exactly what the extremists had sought in their violent attacks."

The legacy of ethnic conflict and polarisation went much beyond the genocide. After Tutsi groups gained power and ended the genocide of Tutsis by Hutus, there were repeated counter-killings undertaken by Tutsi extremists. The large refugee flows destabilised the entire region and contributed to the two Congo Wars (Prunier, 2009). These wars drew several countries of the region into violence and were the deadliest since the Second World War. key figure in the first Congo War was Mobutu, dictator of Zaire. He followed the strategy of "ethnicizing the political situation" (Prunier, 2009) in order to sustain his own power. He purposely destabilised the region by manipulating refugees and constantly changing who is favoured by his regime. Acemoglu, Robinson, and Verdier (2004) describe him as a

clear example of a kleptocratic leader that applied a divide-and-rule strategy to sustain his own power.

Yugoslavia

In the late 1980s, Yugoslavia was a relatively open country that allowed for free travel of people and movement of goods. As Woodward (1995) points out, Yugoslavia was in a good position to make a successful transition to a free market economy and it was moving towards an integration with the West.[9] Compared to other countries in the region in 1989, Yugoslavia enjoyed "relative prosperity, freedom to travel and work abroad, and [a] landscape of multicultural pluralism" (Woodward, 1995). This trend changed abruptly after 1989 and over the following two years, Yugoslavia experienced wars, disintegration, and the creation of new states. War in Croatia led to 20,000 deaths and more than 200,000 refugees. In Bosnia-Herzegovina, it generated 2 million refugees and 70,000 fatalities.

Although intrinsic ethnic differences are often used as an explanation for the eruption of violence in the region, the case of Yugoslavia actually provides yet another example of how political elites *construct* ethnicity in order to divide people and sustain their own power. Referring to the Balkan conflict, Fearon and Laitin (2000) make the observation that violence was used strategically:

> "foment outrage among their own moderates, ethnic leaders will provoke interethnic violence."

An important episode discussed by Woodward (1995) describes how Slobodan Milŏsević polarized politics:

> "The Serbian shift was part of a factional struggle within the Serbian party, which culminated at the eighth party plenum (September 23-24), when Slobodan Milŏsević made a successful move to oust his former patron and friend Ivan Stambolić as president of Serbia and to engineer a coup against its Belgrade party organization, Serbia's most powerful (and liberal). Milŏsević accused Stambolić's crowd of being too lenient on Albanians in Kosovo and of

[9] Unless otherwise noted, the historical exposition below is based on Woodward (1995).

> failing to protect Serbian territorial integrity and Serbs and Montenegrins from forced expulsion."

One interpretation of this description is that Slobodan Milŏsević strategically sought to polarize in order to gain power within his own group. By provoking fear of other groups, he could convince his own people that cooperation with them was not an option. Other nationalist leaders followed similar strategies and also used violence such as targeted assassinations in order to strengthen their position.

Besides using rhetoric and violence to polarize society, the elites also explicitly prevented communication between the groups. Woodward (1995) makes the point that by impeding exchange between groups, the elite could strengthen its position within its own group:

> "[...] the party deprived the country of the one potential antidote: a liberal press and communication across ethnic and national communities to challenge political perceptions, prejudices, and interpretations of events. This failure strengthened the political forces operating within republican arenas by preventing competition and isolated those forces most inclined to fight destructive nationalism. "

The descriptions above show that elites can strategically affect trust in order to divide societies and strengthen their position. This mechanism will be a key feature of the model discussed in the next section.

3.4 The Benchmark Model

In this section, we discuss a static model as a benchmark. This allows us to discuss the basic tradeoffs of the elite between war and peace and the people's decision to revolt. Section 3.5 will then extend the model to a dynamic setting to discuss the evolution of trust over time and how it can be affected by the elite.

3.4.1 Environment

The model is based on Besley and Persson (2011) and consists of two ethnic groups, A and B. One group is the incumbent group, while the other is the opponent group. In this section, we assume that group A initially is the incumbent group. Within each group, there are two types of agents: an elite and the people. The elites do not produce. People in each group can produce at home and also benefit from trade with the other group. The incomes of people in the incumbent and the opponent group through home production are y_I and y_O, respectively, and the income from trade as y_T, which for simplicity is assumed to be the same for both groups. In the following, we always denote the people in the incumbent and the opponent group as I and O, respectively. Moreover, we denote the elite in the incumbent group as E.[10]

There are two types of political systems: autocracy and democracy. In autocracy, the elite of the incumbent group controls the government and makes the following decisions. First, it decides whether to be at war or peace with the opponent group. Second, it sets economic policies: taxes and transfers. The role of the people of the incumbent group is to trade with the other group (only under peace) and to choose whether to revolt against the ruling elite. If a revolution occurs, the regime switches to democracy. In democracy, the elite plays no role anymore, people from both groups together run the government, and the equilibrium is laissez-faire. In other words, we assume away tyranny of the majority in democracy. This means that we avoid the case where one group decides policies and exploits the other group.[11] We use superscript W, P, and R to indicate the possible states of the world, i.e. war, peace and revolution, respectively.

The outcome of trade can be "good", in which case it yields the surplus y_H for both groups, or "bad", in which case the surplus is $y_L < y_H$. The probability of getting the good outcome depends on the type of relation between the two ethnic groups. If the relation is of type "beneficial", the outcome is good with probability q_H and bad with probability $1-q_H$; if it is "harmful", the outcome is bad with probability q_L and good with probability

[10] The elite of the opponent group plays no role in the model, as we will see later, which is why we refer to the incumbent elite simply as the elite and use the index E.

[11] Note that we assume that the two groups are of equal size. However, even in this case a majority decision could favour one group.

$1 - q_L$.[12] Naturally, we assume $q_H > 1 - q_L$, that is to say, if the relation is beneficial, the probability of a good trade outcome is higher than under a "harmful" relation. The agents don't know the type of relation for sure, but hold a common belief about the probability of the relation being beneficial, denoted by p.

We interpret this framework as follows. The type of the relation between two ethnic groups, or equivalently, the likelihood of getting good outcomes from trade, depends on the characteristics of the two groups. For example, if two groups have complementary skills, it is more likely that trade and cooperation between them lead to good outcomes. If the two groups can't communicate well during the trade, for example due to conflicting cultures, then there may be very costly frictions in trade that lead to a low trade surplus. Moreover, we interpret p as trust: the higher is trust in the relation (believing that with high probability the relation is beneficial), the higher is the expected trade surplus. It also implies a larger common interest between the two groups. We therefore view trust as the opposite of ethnic polarization, since the latter emphasizes the conflicting interests across groups.

The timing is shown in Figure 3.1 and it is described in detail below:

1. At the beginning of the period, the political regime is autocracy. The level of trust p is given.

2. I people decide to revolt or not.

   (a) If revolution happens, I people pay the (exogenous) cost f^R and the political system switches to democracy. The elite pays the cost f_E^R and disappears. Then, both groups of people live in the laissez-faire equilibrium and the game ends.

   (b) If there is no revolution, then autocracy survives. The elite stays in government and the game moves to step 3.

3. With a continuing autocracy, the elite decides whether to wage war against O or to retain peace.

[12] Although in our model the expected gains from trade cannot be negative even for the "harmful" relation, this could well be the case in reality. We therefore use the term "harmful", although "not beneficial" would also be an accurate description.

(a) If war occurs, I people pay the cost f^W. There is no trade and the game goes to step 5.

(b) If peace is chosen, I people can trade with O people and the game goes to step 4.

4. Trade: If trade occurs – either in democracy or in peaceful autocracy – then its outcome can either be good (y_H) or bad (y_L) as described in the text above.

5. Political turnover: In autocracy, with probability π^s, the existing incumbent group remains as the incumbent group, and with probability $1-\pi^s$, the other group becomes the new incumbent. The superscript $s \in \{W,P\}$ denotes the state the the world: war or peace. Turnover is less likely when there is war, such that $\pi^P < \pi^W$. In democracy, as both groups of people live in the laissez-faire equilibrium, the distinction of incumbent and opponent group no longer exists.

6. Tax and transfer: In autocracy, the elite from the incumbent group chooses the tax rate on the total income of people, including production and trade surplus. Moreover the elite decides transfer to the people. In democracy, the laissez-faire equilibrium implies that there are no taxes and transfers.

7. Finally, trade outcome realizes and incomes, taxes, and transfers are allocated.

Following Beasley and Persson (2011), we assume that the maximal tax rate is exogenously determined by the state capacity, denoted by τ. The minimal transfer to I people is a θ fraction of total taxes, which is also exogenously determined by checks and balances. Transfers to O people are 0. To understand the state capacity, we can think of the following case: tax payers can hide each unit of income at cost τ. The more capable the state is, the higher costs tax payers have to pay to hide their income. The state, in order to maximize the tax income, prefers to set the tax rate as high as τ, but it can not set it higher, as the Laffer curve goes down towards 0 for tax rates higher than τ. Moreover, the elite sets transfers as low as possible. Therefore, in autocracy, the transfers to I people are a θ fraction of total tax and to O people they are 0. Note that transfers happen after the revolution decision by the people. This implies that a promise by the elite to transfer more than the minimum

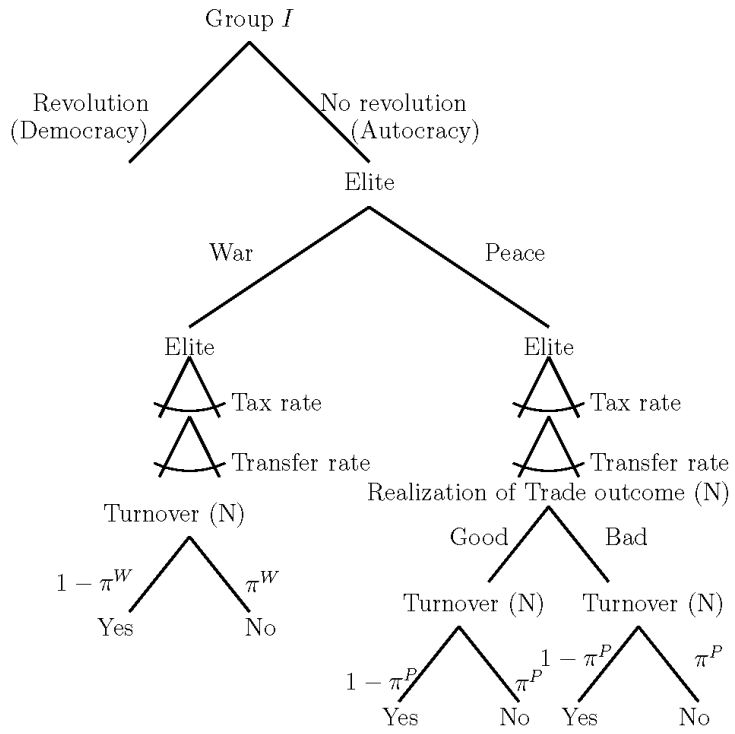

Figure 3.1: Game Tree

The figure shows the timing of the baseline model. (N) denotes nature.

amount given by checks and balances is not credible. Therefore, a promise of transfers cannot be used by the elite to prevent revolution.

3.4.2 Incomes

The equilibrium concept is sub-game perfect Nash equilibrium. We solve the model by backward induction. First of all, we calculate the payoffs of the game, i.e. the incomes of the agents – the elite and the incumbent people – in all the states of the world, i.e. in war, peace, and revolution.

3.4.2.1 I People

The expected trade surplus given belief p is the following:

$$\begin{aligned} E[y_T] &= p(q_H y_H + (1-q_H) y_L) + (1-p)(q_L y_L + (1-q_L) y_H) \\ &= p(q_H + q_L - 1)(y_H - y_L) + (q_L y_L + (1-q_L) y_H). \end{aligned}$$

Recall that $q_H > 1 - q_L$. We know that y_P is positively correlated with p, namely, the higher trust, the higher is the expected trade surplus. To simplify the expression, we set

$$\begin{aligned} y_H &= \frac{q_L}{q_H + q_L - 1}, \\ y_L &= \frac{q_L - 1}{q_H + q_L - 1}, \end{aligned}$$

which leads to

$$E[y_T] = p.$$

The key relationship between the trade surplus and the belief of the other group is preserved, so this normalization is without loss of generality and does not change the results of this model.

The final expected income of I people includes production, trade surplus, and transfers minus taxes. I people get positive transfers if group I stays the incumbent group. If a political turnover occurs, I people become the opponent group and get 0 in transfers. This implies the following incomes of the I people in each state of the world:

- Peace: $y_I^P = (1-\tau)(y_I + p) + \pi^P \theta \tau (y_I + y_O + 2p)$.
- War: $y_I^W = (1-\tau) y_I + \pi^W \theta \tau (y_I + y_O - f^W)$.
- Revolution: $y_I^R = y_I + p - f^R$.

We assume that $\pi^W > \pi^P$, implying that for group I, there is a tradeoff between peace and war: in peace, there is extra income from trade, while in war, the probability of staying as the incumbent group is higher. Peace dominates war for I people if and only if trust, or equivalently, the expected gain from trade, p, is large enough. In this case, $y_I^P \geq y_I^W$, which

is equivalent to the following:

$$\begin{aligned}(1-\tau)(y_I+p)+\pi^P\theta\tau(y_I+y_O+2p) &\geq (1-\tau)y_I+\pi^W\theta\tau(y_I+y_O-f^W) \iff \\ p &\geq \frac{\left(\pi^W-\pi^P\right)\theta\tau(y_I+y_O)-\pi^W\theta\tau f^W}{1-\tau+2\pi^P\theta\tau} \\ &\doteq p^W.\end{aligned}$$

In other words, if $p < p^W$, I people strictly prefer to go to war instead of peace. When do I people prefer revolution to peace? Under the condition that $\theta < \frac{1}{2\pi^P}$, we can derive the threshold for revolution:

$$\begin{aligned}y_I^R &\geq y_I^P \iff \\ y_I+p-f^R &\geq (1-\tau)(y_I+p)+\pi^P\theta\tau(y_I+y_O+2p) \iff \\ p &\geq \frac{\pi^P\theta\tau y_O-\left(1-\pi^P\theta\right)\tau y_I+f^R}{\tau(1-2\pi^P\theta)} \\ &\doteq p^R.\end{aligned}$$

The condition that $\theta < \frac{1}{2\pi^P}$ implies that the transfer to I people can not be too large. If it is violated, I people expect to get more transfers than what they pay in taxes. Their income would then increase by more than 1 unit when trust and the trade surplus for them increase by 1 unit. This would lead to the case that I and O people cooperate to start the revolution only when trust between them is very low. This case is unlikely to happen in the real world and we therefore rule it out. We maintain this as an assumption throughout the paper:

Assumption 3.1. $\theta < \frac{1}{2\pi^P}$.

3.4.2.2 Elite

If it can maintain its power, then the final income of the elite consists of taxes minus transfers. If there is a revolution, then it incurs a high cost because it is thrown out of the government. The payoffs for each state of the world are therefore as follows:

- Peace: $y_E^P = \pi^P(1-\theta)\tau(y_I+y_O+2p)$.
- War: $y_E^W = \pi^W(1-\theta)\tau(y_I+y_O-f^W)$.

- Revolution: $y_E^R = -f_E^R$.

If there is no revolution, then the elite prefers peace to war when p is large enough:

$$\begin{aligned} y_E^P &\geq y_E^W \iff \\ \pi^P(1-\theta)\tau(y_I+y_O+2p) &\geq \pi^W(1-\theta)\tau(y_I+y_O-f^W) \iff \\ p &\geq \frac{\left(\pi^W-\pi^P\right)\tau(y_I+y_O)-\pi^W\tau f^W}{2\pi^P\tau} \end{aligned}$$

It can be verified that $p_E^W > p^W$ by observing the following:

$$\begin{aligned} p^W &= \frac{\left(\pi^W-\pi^P\right)\theta\tau(y_I+y_O)-\pi^W\theta\tau f^W}{1-\tau+2\pi^P\theta\tau} \\ &= \frac{\left(\pi^W-\pi^P\right)\tau(y_I+y_O)-\pi^W\tau f^W}{\frac{1-\tau}{\theta}+2\pi^P\tau} \\ &< \frac{\left(\pi^W-\pi^P\right)\tau(y_I+y_O)-\pi^W\tau f^W}{2\pi^P\tau} \\ &\doteq p_E^W. \end{aligned}$$

willing to go to war when the trust is relatively low. The reason is that if the political turnover occurs, the elite loses more compared to I people, who can still get the after-tax income even under the rule of the other group. This difference between the elite and the people in the willingness to go to war is frequently discussed in the literature. For example, Rohner, Thoenig, and Zilibotti (2013a) use such an argument to generate a random war that is out of the control of people. Jackson and Morelli (2007) argue that if the political process is captured by a biased political elite, then the war occurs against the interest of people. Here we offer an explanation for why the elite is more eager to go to war than the people. One reason is the fear of political turnover. But there is another reason, which is the fear of revolution when trust is high. This second reason will be discussed in section 3.5 where we present the dynamic model.

3.4.3 Equilibrium Outcomes

Given the incomes of agents, we can discuss the choices of the elite and the I people for different levels of trust. This is summarized formally in the following proposition.

Proposition 3.2. *If trust is sufficiently low ($p \leq p^W$), both the elite and the incumbent people prefer war over peace. The elite is more eager to go to war than the people. Namely, if trust is in the interval $\left[p^W, p_E^W\right]$, then the elite prefers war to peace while incumbent people prefer peace. Moreover, if trust is high enough, i.e., $p \geq p^R$, incumbent people prefer revolution to peace.*

The cut-off values of trust, p^W, p_E^W, and p^R, can be ranked from low to high given certain values of the cost parameters f^R and f_E^R. In the case where f^R and f_E^R are sufficiently large, we have $p^W < p_E^W < p^R$, and the elite never prefers revolution to peace. By sufficiently large costs, we mean $f^R > \underline{f^R}$ and $f_E^R > \underline{f_E^R}$, where the expressions for $\underline{f^R}$and $\underline{f_E^R}$ are given in the appendix. Then, the equilibrium outcomes for all p can be characterized. We summarize this in the next proposition (proof in the appendix) and it is illustrated in Figure 3.2.

Proposition 3.3. *Given that $f^R > \underline{f^R}$ and $f_E^R > \underline{f_E^R}$, we have $p^W < p_E^W < p^R$, and the equilibrium outcomes for different levels of p (from low to high) are given by the following four cases.*

$p < p^W$: war. Both elite and incumbent people prefer war, since the expected trade surplus under peace is too low.

$p^W < p < p_E^W$: war. Though incumbent people prefer peace given the large trade surplus, the elite decides to stay at war. The elite has more to lose if there is a political turnover.

$p_E^W < p < p^R$: peace. There is enough trade surplus to gain and no threat of revolution. The elite prefers peace and so do incumbent people.

$p > p^R$: revolution. The expected trade surplus is so large that incumbent people decide to revolt. The elite is overthrown and democracy is established. The elite never prefers revolution.

If p is at the threshold level $p = p_E^W$, then the equilibrium outcome can be either war or peace, because the elite is indifferent between the two states and pure strategies and mixed

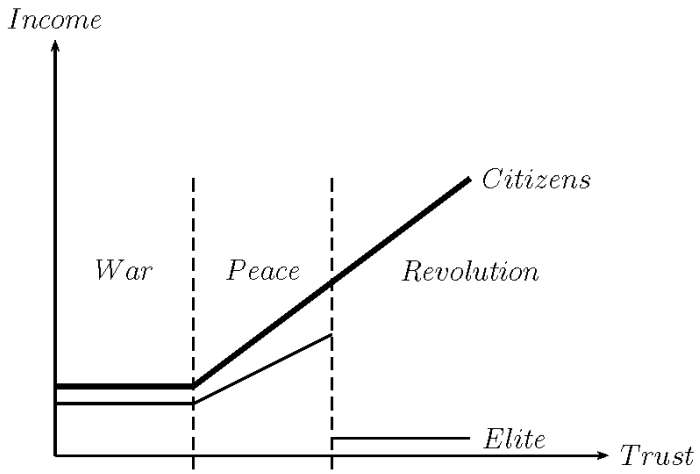

Figure 3.2: Trust and Income
The figure shows the relationship between trust and income. The thin solid lines refers to the elite, while the thick solid line represents the citizens. The dashed lines separate the different states.

strategies between war and peace can all be equilibria. The analogous result holds for the case $p = p^R$.

The revolution occurs when trust is too high. This implies that the elite may have an incentive to prevent trust from getting too high. Starting a conflict and thus preventing the updating of beliefs during trade interactions gives the elite a way of limiting trust. The dynamic model in section 3.5 allows discussing how the elite can apply this divide-and-rule strategy.

3.5 The Dynamic Model

In this section, we consider a dynamic model that includes the evolution of trust. With this model, we study under which circumstances the elite wants to decrease the trust between the two ethnic groups, or equivalently, polarize the society. The elite seeks to influence trust because trust affects the decisions of people to revolt, which in turn influence the income of the elite.

3.5.1 Environment

We assume that the elite lives for infinite periods and is forward-looking. Each cohort of people lives for only one period and in the next period a new cohort of people are born. The assumption that people care only about the current period income simplifies the analysis substantially. Although people's decision on revolution or peace will be different from the current setting if we were to assume that people are forward-looking, we expect the key results of the model to still hold – as long as they still prefer revolution to peace when trust is sufficiently high. We show after the presentation of the dynamic model that this is plausible.

Each period of the dynamic model includes everything that happens in the static model. Moreover, the evolution of the political system and trust over the periods follows the rules below. If democratization occurs in one period, democracy is set up and consolidated for all the periods afterwards. The equilibrium in all these periods will be laissez faire, such that people of both groups simply gain their home production income and the trade surplus. If autocracy survives this period, the next period starts with autocracy.

The dynamics of trust follow the following rule. First, in the beginning of each period, trust is p, i.e., the common belief about the probability that the relation between the two groups is "beneficial". Then, if trade occurs in this period, agents update their belief about the relation after observing the trade outcome, i.e. they update their belief of trust using Bayes' rule. If there is no trade, agents learn nothing about the type of relation. Finally, at the end of each period, there is some probability that the type of the relation switches. The stochastic types are driven by cultural shifts, which can be related to social structure, population mixture, and so on (see Tabellini, 2008, and Rohner, Thoenig, and Zilibotti, 2013a). For example, consider a case where in the beginning both groups consist of mainly young people. Their skills are not complementary with each other and the interactions and trade between young people are full of frictions. Some time later, the population of one group becomes older due to different demographic trends, while the other group stays the same. Now their skills are complementary, since the old group brings in experience and knowledge while the young group contributes for example creativity and energy. The relation between the two groups thereby becomes beneficial and trade and cooperation are

more likely to generate good outcomes. More specifically, we assume that the type of the relation follows a two-state first-order stochastic Markov process with the following transition matrix:

	t B	H
$t-1$		
B	$1-\psi$	ψ
H	ϕ	$1-\phi$

where B and H denote the types of the relation being "beneficial" and "harmful", respectively, and ψ and ϕ are the probabilities that the type changes, conditional on the relation being "beneficial" or "harmful", respectively. Agents are aware of the possibility that types can switch. This implies that agents change their belief even if there was no trade, because they anticipate that the nature of the relation may have switched.

How does trust change when there is a political turnover? We assume that turnover doesn't affect trust. This is a fairly natural assumption in our setting. Trust is the common belief about the probability of the relation being beneficial, based on the common observations of past trade outcomes. It should therefore be the same for both ethnic groups, either as incumbent or opponent.

We also assume that after the turnover, the elite of the old incumbent group loses its rents and exits the economy. If some periods later, this group becomes the incumbent group again, then a new cohort of elite is ruling. Therefore, the current elite's continuation value only comes from the consecutive periods that it stays in power without turnover.

3.5.2 Solution

In the beginning of each period, we still denote the variables of the incumbent group with subscript I, and the variables of the opponent group with subscript O. If at the end of this period the incumbent group becomes opponent, we still keep the subscript I to denote the (new) incumbent, but keep in mind that the values of the variables changes. For instance, in the period when group A is the incumbent group, y_I takes the value of the home-production income of people A, while in the next period when group B is the incumbent group, y_I' takes the value of the home-production income of people B, which can be different from y_I.

The incomes and the choices of people are exactly the same as in the static model as each cohort lives for only one period. Therefore, if $p > p^R$, I people choose revolution, otherwise the elite stays in power and sets government policies. Essentially, this is what we need to know about the people. This is why, if people were to live for infinite periods, we would expect the results of the model not to be too different – as long as people choose to revolt when trust is very high.

We now show how trust evolves. Lets say that initially trust is at some level p at the beginning of the period and that trade takes place. Then, all agents update their belief according to Bayes' rule based on the trade outcome, either in democracy or in autocracy with peace, i.e.

$$p^+ = \begin{cases} \frac{pq_H}{pq_H+(1-p)(1-q_L)} & \text{if } y_T = y_H, \\ \frac{p(1-q_H)}{p(1-q_H)+(1-p)q_L} & \text{if } y_T = y_L. \end{cases}$$

If there is no trade due to the war, then there is no new information about the type of the relation, so the belief stays the same at this moment, i.e. $p^+ = p$. But at the end of the period, the stochastic shock to the type of the relation realizes. After that, the belief that the type is "beneficial" becomes $p' = p^+(1-\psi) + (1-p^+)\phi = p^+(1-\psi-\phi) + \phi$, and this is also the level of trust at the beginning of the next period. The evolution of trust across two periods is summarized below:

$$p' = \begin{cases} (1-\psi-\phi)\, p + \phi & \text{if } s = W, \\ (1-\psi-\phi)\, \frac{pq_H}{pq_H+(1-p)(1-q_L)} + \phi & \text{if } s = P, y_T = y_H, \\ (1-\psi-\phi)\, \frac{p(1-q_H)}{p(1-q_H)+(1-p)q_L} + \phi & \text{if } s = P, y_T = y_L. \end{cases}$$

We can see that if war continues for infinitely many periods, trust converges to the "natural" level $\frac{\phi}{\phi+\psi}$ with the auto-correlation $(1-\psi-\phi)$. We want the probability of the change of types not to be too high and the "natural" level of trust not to be too low nor too high, such that people prefer peace at this level. therefore maintain the following assumption:

Assumption 3.4. *$\psi \leq \frac{1}{2}$, $\phi \leq \frac{1}{2}$ and $\frac{\phi}{\phi+\psi} \in \left[p^W, p^R\right]$.*

If revolution happens, then the equilibrium outcomes are just a repetition of the case of revolution in the static model. If there is no revolution, then in autocracy the elite's choice

sets are the same as in the static model: war or peace, tax and transfer. Tax and transfer are trivially determined by the maximal tax rate and the minimal transfer, as in the static model. However, when the elite decides on war or peace, it now maximizes the expected life-time income, given the current period's level of trust:

$$V(p) = \begin{cases} \max_{s\in\{W,P\}} y_E^s(p) + \beta \pi^s E^s V(p'), & \text{if } p \leq p^R, \\ -f_E^R, & \text{if } p > p^R, \end{cases}$$

where y_E^s is the elite's current period income, π^s is the probability that the current incumbent group stays in power (both depending on $s \in \{W,P\}$), p' is the level of trust in the next period, and β is the discount factor. From this follows the law of motion discussed above.

We can characterize important properties of the value function and of its solution, given certain conditions and a value of p. We can for example consider the cases when (1) both the cost of war and trust are low, or (2) when both the cost of revolution for the elite and trust are high. discuss the elite's behavior and the equilibrium outcomes in each of these cases.

First, when trust and the cost of war are both low enough, then there is not much to gain from peace and trade, and the loss from going to war is small. Moreover, war implies a lower chance of political turnover compared to peace, which increases the elite's expected income both this period and in the future. This property can be summarized in the following proposition (proof in the appendix).

Proposition 3.5. *Suppose that $f^W < \frac{\pi^W - \pi^P}{\pi^W}(y_I + y_O)$, then when p is sufficiently low, the elite wages war. More precisely, for $p = 0$, the elite chooses war; and if $V(p)$ is continuous at $p = \phi$, then there exist a $p_E^W > 0$, such that for all $p < p_E^W$, war is chosen as well.*

Second, consider the following scenario: in a given period, trust is relatively high, peace is chosen by the elite, trade occurs between the two groups for many periods, and there are enough good trade outcomes such that trust keeps increasing. The increasing trust raises the trade surplus and income of people in peace, but is it good news for the elite? Not necessary. While the high level of trust implies gains from trade and thus large resources in the society that the elite can potentially extract, it is also a threat to the elite, since a too high level of trust may trigger a revolution. In particular, when $p > p^R$, I

people find that cooperating with O people in democracy (where they are not taxed by an elite) is better than autocracy (in which they extract income from the O people but are also being expropriated by their own elite). If p is large enough and very close to p^R, choosing peace means that with some positive probability the trade outcome will be good. Then, p may increase to a level higher than p^R, which will trigger a revolution in the next period. Since revolution implies a high cost for the elite, it can be optimal for the elite to choose war instead of peace even when trust is already high, since during the war the trust is expected to regress back to the natural level p_0 where there is no risk of revolution. To avoid that trust can grow above this critical level after sufficiently many good trade outcomes, the elite can strategically start a war to prevent people from seeing good trade outcomes. In the appendix, we derive the "sufficiently large" cost of revolution as $\underline{f_E^R} \doteq \frac{\pi^P(1-\theta)\tau(y_I+y_O+2p^R)-\pi^W(1-\theta)\tau(y_I+y_O-f^W)+\beta\pi^P\left(1-q^R\right)\bar{V}-\beta\pi^W\underline{V}}{\beta\pi^P q^R}$. Then we have the following proposition (proof in the appendix):

Proposition 3.6. *Suppose that $f_E^R > \underline{f_E^R}$, then if p is sufficiently large, the elite chooses to wage war, while people would prefer peace. The war not just prevents trust from growing, but it leads to declining trust. The elite increases polarization and reduces the interaction between the two groups in order to increase its own probability of staying in power. More precisely, when $p = p^R$, then the elite prefers war over peace; and there exists a $p_E^R < p^R$, such that for all $p > p_E^R$, the elite chooses war.*

In this case, war occurs when trust is high and when there is a lot to gain from trading in peace. This is an interesting result, since high trust makes peace more attractive for the society overall. The reason for this result is that the elite's interest is different from the interest of the society and of the people. It is true that the higher level of trust generates larger total income in the whole society, but it also changes the allocation of income by triggering a revolution, which reduces the income of the elite drastically. This is the second reason why the elite decides to go to war and to stop the trade interaction against the interest of people: the fear of an increase in trust and of the revolution that this may trigger.[13]

If war starts and trade is interrupted at a high level of trust, then there is no chance of a further increase in trust. Although the termination of trade between the two ethnic groups

[13]Recall from the discussion of the static model that the other reason why the elite is more likely to go to war than the people is that they have more to lose from a transition of power to the other group.

does not directly harm trust, it in fact prevents the potential growth of trust. Trust during war gradually converges back to its natural level which is lower than the threshold where a revolution can occur. This means that after the elite purposely stops trade, polarization can increase. This strengthens the power of the elite and prevents revolution. This is one of the key findings of this paper and it represents the divide-and-rule strategy that has frequently been described as a source of ethnic tensions.

Given the conditions in the propositions above, the dynamic model generates certain results that are similar to the ones in the static model in the following respects. First, if the cost of war is sufficiently small, the elite chooses to go to war when trust is low, because the benefit from peace, i.e., the trade surplus, is expected to be low. The incumbent group fights with the opponent group in order to stay in power with higher probability. Second, if trust is too high, I people prefer revolution and democracy together with O people, in which they benefit from the expected high trade surplus and are no longer expropriated by the elite. Such a revolution is very costly for the elite since they lose all their rents. While in the static model trust is exogenously given, the dynamic model allows us to analyze how trust evolves. More importantly, the dynamic model allows us to see how the elite strategically influences trust. The elite achieves this by going to war in order to stop further increase in trust, which otherwise would lead to revolution.

How does trust evolve over time? If trust is very low, then the elite (and also the incumbent people) prefer to be at war with the other group. In the absence of trade interactions during the war, trust regresses to the natural level due to the stochastic switch of types. After some periods, trust may reach a level that can lead to peace. Then, if there are sufficiently many good trade outcomes during peace, trust rises further. But if trust becomes so high that the elite sees the potential of a revolution in the near future, then it wages a war again to stop trade. During the war there are again no trade interaction. This war, in a situation when the level of trust is high, is at the cost of the people and the society. But it helps the elite to stay in power, since it keeps trust low (and thus polarization high), so that the incumbent people do not have sufficient common interest with the other group to revolt against the own elite. During the war, trust may regress down due to the stochastic relation type such that there is no threat of revolution anymore. Then, the elite may stop the war and allow trade again.

This model helps us to understand some of the most disturbing episodes of violence in recent history. As we have shown in Section 3.3, elites in Rwanda and Yugoslavia have purposely destroyed the common interest of people in order to maintain or increase their power. This destruction of the common interest often took the form of ethnic violence which hampered trust. The lack of trust then triggered further violence between groups who thought that they have little to gain from peaceful interaction. Our model provides an explanation for why rational people follow leaders into destructive violence. Some previous approaches to answering this question have assumed that elites can directly manipulate people's beliefs (for example Posner, Spier, and Vermeule, 2010). In contrast, agents in our model are perfectly aware of other agents strategies. The elite's ability to initiate conflict to prevent economic interaction allows it to induce people to polarise by affecting trust between people of different groups.

3.6 Conclusion

We propose a theory of divide-and-rule where political elites strategically initiate conflicts between ethnic groups in order to polarize society and thus sustain their own power. We model polarization as a lack of trust that is shaped by trade interactions between agents of the different ethnic groups. The elite can prevent trust from emerging by starting a conflict that interrupts trade. The elite follows this strategy when there is a threat of revolution. The model also generates that an elite is more likely to seek war than the people. This is the case because the elite has the double fears of losing power to the other group and to be overthrown by the own group if the common interest between groups becomes too large.

We document that our model is consistent with a number of cases of large-scale ethnic violence, in particular with the incidents in Rwanda and Yugoslavia. These cases show that polarization was not simply exogenously given, but to some extent *constructed* by power-seeking elites. Violence often had the goal of destroying trust and creating instability. This allowed elites to exercise their power more freely. An implication of these observations is that treating conflicts as "inherently ethnic" may be misleading. Our model shows the elite's role in creating ethnic polarization and it therefore enables us to discuss possible counter-measures. Generally speaking, attempts to prevent ethnic violence should

pay considerable attention to the role of the elite. Promising ways to achieve this would be to reduce elites' incentive or ability to reduce trust. This is especially important in cases where the elite may attempt to do so because they are threatened by a *high* level of trust. In such cases, the interests of the elite diverge from those of society overall, since the elite values being in power more than possible gains from trade.

Appendix A

Appendix of Chapter 1

A.1 Details on the Equilibrium Given Capital Allocations

In subsection 1.3.2, we state the equilibrium outcomes in democracy and oligarchy without going into details. Most importantly, we state that to sustain oligarchy, the political constraint is that there must be sufficiently many supporters for oligarchy, and a worker supports oligarchy if she gets higher income than in democracy. In facts, those statements can be derived from a sequential game played between workers and elites, as follows.

In democracy, the majority, namely workers, elect a representative worker into the government, then she decides government policies to maximize her income, or equivalently, workers' income. The government policies in democracy are the following: the government taxes elites and entrepreneurs, transfers the tax income to workers and it doesn't regulate the S sector wage. First, taxing elites and entrepreneurs involves no cost and making transfer to them gives no benefits, so they are taxed and get no transfer. Second, since tax income is transferred to workers, taxing workers and transferring this income back to them is equivalent to no tax on them. Third, the government has no incentive to distort the wage in S sector and the labor market allocation. Setting a binding minimal wage in S sector leads to inefficient labor allocation, lower total output and lower final income for workers. So the government does not regulate wage in democracy. The labor market is competitive and efficient.

Given the above analysis, the timing of events is the following:

1. Capital in S and P sectors are given as K_S, K_P.
2. Workers decide tax rates on elites and entrepreneurs.
3. S and P firms hire workers in the competitive labor market and produce. S and P firms produce. Capital incomes and wages are distributed.
4. Elites and entrepreneurs decide whether to hide income at the cost $\bar{\tau}$.
5. Tax are collected and and transferred to workers.

Now we can solve the equilibrium in democracy from backwards. In stage (4), tax payers choose to pay the tax if $\tau \leq \bar{\tau}$, otherwise they hide the income [1]. In stage (3), competitive labor market implies that wage in S and P firms are the same and equal to the marginal productivity of labor:

$$w^D = (1-\alpha)(z_S K_S)^{\alpha} \left(L_S^D\right)^{-\alpha} = (1-\alpha)(K_P)^{\alpha} \left(L_P^D\right)^{-\alpha}.$$

The market clearing condition $L_S^D + L_P^D = 1$ helps us to pin down the labor allocation:

$$\begin{aligned} L_S^D &= \frac{z_S K_S}{z_S K_S + K_P}, \\ L_P^D &= \frac{K_P}{z_S K_S + K_P}. \end{aligned}$$

The pre-tax income for elites and entrepreneurs are the capital returns, respectively:

$$\begin{aligned} \pi_S^D &= \alpha (z_S K_S)^{\alpha} \left(L_S^D\right)^{1-\alpha}, \\ \pi_P^D &= \alpha (K_P)^{\alpha} \left(L_P^D\right)^{1-\alpha}. \end{aligned}$$

In stage (2), the government decides $\tau = \bar{\tau}$ to get the maximum tax income, given the tax-payers' strategy on income hiding, and the fact that in this static model without investment, tax is not distortive. Now we have solved the problem and we can write the final income of

[1] Without loss of generality, we assume that when tax payers are indifferent between paying tax or hide the income, they pay the tax.

workers, elites and entrepreneurs as follows:

$$\begin{aligned} y_w^D &= w^D + \bar{\tau}\left(\pi_S^D + \pi_P^D\right) \\ &= \left(1 + \bar{\tau}\frac{\alpha}{1-\alpha}\right) w^D, \\ y_e^D &= (1-\bar{\tau})\,\pi_S^D, \\ y_p^D &= (1-\bar{\tau})\,\pi_P^D. \end{aligned}$$

The transfer to workers is $\bar{\tau}\frac{\alpha}{1-\alpha}w^D$ simply because the tax base - capital income - is $\frac{\alpha}{1-\alpha}$ times labor income.

To sum up, in democracy, no distortive policy and competitive labor market imply first best allocation. elites and entrepreneurs get $(1-\bar{\tau})$ fraction of capital income, respectively. The final income of workers is $\bar{\tau}\frac{\alpha}{1-\alpha}$ times larger than their wage income.

In oligarchy, elites control the government to maximize their income. In this context, we can use the government and the elites interchangeably. First, the government decides to democratize or not. If yes, the economy becomes exactly the same as in democracy. Otherwise, the oligarchic government decides the two policies - tax and state sector wage regulation. Given the policies, agents know their final incomes in oligarchy. Comparing incomes in oligarchy and in democracy, an agent decides to support oligarchy or not. If and only if large enough fraction of citizens choose to support the oligarchy, the regime survives. Otherwise democratization occurs. In the latter case, the economy is again the same as in democracy. The minimal size of supporters is denoted as $\underline{L}$, which is exogenous. Because the population of elites and entrepreneurs is as small as measure 0, oligarchy is sustained if and only if more than $\underline{L}$ workers support it [2].

[2]The above setting, that political support decides political regime, is based on the framework of Acemoglu et al. (2012). In their language, a coalition of some agents is assigned with a level of political power, and if the power is large enough, the coalition can change the political system according to their preference. In our context, in oligarchy, elites as the ruling group are granted some level of political power, denoted as ω_e. Each worker has political power ω_w, and each entrepreneur has ω_p. The aggregate political power of entrepreneurs is just 0 given its small size. Workers can change the political regime from oligarchy to democracy if and only if they form a coalition of size L_r and $\frac{\omega_w L_r}{\omega_w+\omega_e} > \alpha \Longleftrightarrow L_r > \alpha\frac{\omega_w+\omega_e}{\omega_w}$, where α is exogenous. In other words, the sustain a oligarchy, there must be at least $1-\alpha\frac{\omega_w+\omega_e}{\omega_w}$ workers supporting elites. We denote this size as $\underline{L}$.

The government taxes private entrepreneurs and P workers, but doesn't tax elites and S workers. The logic is the following. Taxing entrepreneurs involves no cost for elites. Tax on elites will be transferred back to themselves, so it is equivalent to no tax. Tax on S workers gives elites no benefit but implies larger distortions, because the government anyway needs to increase S workers' after-tax income to the level in democracy if the government wants to sustain oligarchy. It implies that the tax has to be compensated by higher wage payment to S workers, so tax on S workers should be 0. Furthermore, there is no cost to tax P workers, so the tax rate on P workers can be positive. Tax rate on entrepreneurs is also not determined yet.

Given the tax system, the timing of events in oligarchy can be simplified as follows:

1. Elites choose to democratize or not. If yes, the economy is the same as in democracy. otherwise the following events occurs.

2. Elites sets minimal wage in S sector as w_S.

3. *S* firms and P firms hire workers.

4. Workers in *S* and *P* sectors decide to support the current regime or democratization, simultaneously.
   (a) If more than $\underline{L}$ workers support the regime, oligarchy is sustained.
   (b) If fewer than $\underline{L}$ workers support the regime, democratization occurs.

5. If oligarchy is sustained, *S* and *P* firms produce. Capital income, state and private sector wages are distributed accordingly.

6. Tax is collected and transferred to elites.

Now we can solve the rest of the problem from backwards. First, in stage (5), firms' labor demand gives the relation between wage and labor. S firms, given the regulated wage w_S, choose labor demand L_S so that wage equals marginal productivity. Similar for private firms who face market wage w_P.

$$w_S = (1-\alpha)(z_S K_S)^{\alpha} L_S^{-\alpha}, \tag{A.1}$$

$$w_P = (1-\alpha) K_P^{\alpha} L_P^{-\alpha}. \tag{A.2}$$

Then state and private workers' final incomes are the following:

$$\begin{aligned} y_{wS} &= w_S, \\ y_{wP} &= (1-\tau)\, w_P. \end{aligned}$$

Second, we look at workers' political decisions in stage (4). No matter what others choose, a weakly dominant strategy is to support oligarchy if and only if the income is higher than in democracy. Without loss of generality, we assume that workers use the this strategy[3]. In other words, a worker's political choice truthfully reflects her economic interest. S and P workers support oligarchy if and only if

$$\begin{aligned} y_{wS} &\geq y_w^D, \\ y_{wP} &\geq y_w^D, \end{aligned}$$

respectively.

Then the rest is the same as the discussion in subsection 1.3.2 of the body part: in stage 2, w_S is set to be high enough to guarantee support from S workers: $w_S \geq y_w^D$, while the general equilibrium effect reduces the P worker wage $w_P \leq w^D < y_w^D$, and private workers never support oligarchy. When S sector capital is large enough, more than $\underline{L}$ workers are hired by state firms, and oligarchy is sustained.

A.2 Proof of Proposition 1.4

There are two possible cases that makes whether oligarchy can be sustained in this period.

1. $[\underline{L}, \bar{L}] = \emptyset$. Oligarchy can not be sustained. The government can not strategically make sufficient workers support the oligarchy: if it sets $L_S < \underline{L}$, there's not sufficient support; but if $L_S \geq \underline{L} > \bar{L}$, workers in S sector don't get high enough wage. The only possible political outcome is democracy. Since $\underline{L}$ is exogenous and $\bar{L} = \nu L_S^D =$

[3]Other strategies and equilibria give the same outcome. For example, giving support to oligarchy when income in democracy is higher is a best response if and only if this worker's choice doesn't change the outcome. So a worker may use this different strategy, but the equilibrium outcome is still the same.

$\nu \frac{zK_S}{zK_S+K_P}$ depends on $\frac{zK_S}{K_P} \doteq k$, we can simplify the condition $\bar{L} < \underline{L}$ to $\nu L_S^D = \nu \frac{k}{1+k} < \underline{L} \Leftrightarrow k < \underline{k} = \begin{cases} \frac{\underline{L}}{\nu - \underline{L}} & \text{if } \nu - \underline{L} > 0 \\ +\infty & \text{if } \nu - \underline{L} \leq 0 \end{cases}$, given $k > 0$. In the case $\underline{k} = \frac{\underline{L}}{\nu - \underline{L}}$, we know that when there are not enough capital in S sector compared to the capital in P sector, oligarchy can't be sustained.

2. $[\underline{L}, \bar{L}] \neq \emptyset$. Oligarchy can be sustained. The government can choose to sustain the oligarchy by setting some $L_S \in [\underline{L}, \bar{L}]$ or choose democratization.

Appendix B

Appendix of Chapter 2

B.1 Technical analysis and extensions related to Section 2.2

Proof. [Proof of Proposition 2.1] The strategy of the proof is to verify that the pension policy (2.8)-(2.9) implies individual consumption and labor supply consistent with the first best allocation (2.3)-(2.5). The pension policy, (2.8)-(2.9) and zero taxes, yields

$$\begin{aligned} c_t^Y &= (1+\beta)^{-\frac{1}{1+\theta}} \left((\phi R)^{t(1+\theta)} \left(\frac{1+\theta}{\lambda w_t} \right)^{1+\theta} \right)^{\frac{1}{1+\theta}} w_t \\ &= \frac{1+\theta}{\lambda} (1+\beta)^{-\frac{1}{1+\theta}} (\phi R)^t \\ h_t^{\frac{1}{\theta}} &= w_t \frac{\lambda}{1+\theta} (1+\beta)^{\frac{1}{1+\theta}} (\phi R)^{-t}. \end{aligned}$$

The implied between-generation consumption growth is then, for $t \geq 1$

$$\frac{c_{t+1}^Y}{c_t^Y} = \frac{\frac{1+\theta}{\lambda} (1+\beta)^{-\frac{1}{1+\theta}} (\phi R)^{t+1}}{\frac{1+\theta}{\lambda} (1+\beta)^{-\frac{1}{1+\theta}} (\phi R)^t} = \phi R,$$

as in the first best allocation (2.3). Finally, we show that the Ramsey allocation yields the same intratemporal first-order condition as the first best allocation, i.e., $h_t^\theta \cdot c_t^Y / w_t = 1$:

$$\begin{aligned} & (h_t)^\theta \cdot c_t^Y \frac{1}{w_t} \\ = \; & w_t \frac{\lambda}{1+\theta} (1+\beta)^{\frac{1}{1+\theta}} (\phi R)^{-t} \cdot \frac{1+\theta}{\lambda} (1+\beta)^{-\frac{1}{1+\theta}} (\phi R)^t \frac{1}{w_t} = 1. \end{aligned}$$

[Proof of Proposition 2.3] The planner's first-order conditions for consumption are the same as in the first-best allocation (2.3). The planner's first-order conditions for labor supply are modified to

$$h_t^{\frac{1}{\theta}} = \frac{w_1}{1+g}\lambda - \mu \quad \text{for } t \geq 1 \tag{B.1}$$

$$h_0^{\frac{1}{\theta}} = w_0\lambda - \mu_0, \tag{B.2}$$

where μ_0 and μ are Lagrange multipliers on the non-negative-pensions constraint. Note that since equation (B.1) is the same for all t, μ and h_t must be constant for all $t \geq 1$. Since $\lambda \leq (1+\beta)^{1/(1+\theta)} / w_0$, equation (2.12) implies that

$$\zeta_0 \geq R\left(\left(\left(w_0 (1+\beta)^{1/(1+\theta)} / w_0\right)\right)^{-(1+\theta)} (1+\beta) - 1\right) = 0.$$

The Kuhn-Tucker condition then implies that $\mu_0 = 0$ so that (B.2) is satisfied. Setting $\tau_0 = \mu_0 = 0$, the individual's optimal consumption is then determined by the budget constraint and equation (2.7), given by

$$\begin{aligned} c_0^Y &= \frac{1}{(1+\beta)}\left(w_0 h_0 + \frac{\zeta_0 w_0 \bar{h}_0}{R}\right) = \frac{1}{(1+\beta)}\left(1+\frac{\zeta_0}{R}\right) w_0 h_0 \\ &= \frac{1}{(1+\beta)}\left(1+\frac{\zeta_0}{R}\right) w_0 (1+\beta)^{\frac{\theta}{1+\theta}} \left(1+\frac{\zeta_0}{R}\right)^{-\frac{\theta}{1+\theta}} \\ &= \frac{1}{(1+\beta)} w_0 (1+\beta)^{\frac{\theta}{1+\theta}} \left(1 + \frac{R\left((\lambda w_0)^{-(1+\theta)} (1+\beta) - 1\right)}{R}\right)^{1-\frac{\theta}{1+\theta}} = \frac{1}{\lambda}, \end{aligned}$$

which is consistent with λ being the planner's marginal utility of c_0^Y. The planner's intra-temporal first-order condition then holds: $(h_0)^{1/\theta} = w_0/c_0^Y$.
Since $\zeta_t = 0$ for $t \geq 1$, labor supply is, according to (2.7), given by $h_t = (1+\beta)^{\theta/(1+\theta)}$. Since by assumption $\lambda \geq (1+\beta)^{1/(1+\theta)}(1+g)/w_1$, it follows that

$$h_t = (1+\beta)^{\frac{\theta}{1+\theta}} < \left(\frac{w_1}{1+g}\lambda\right)^{\theta}.$$

Equation (B.1) therefore holds, confirming that $\zeta_t = 0$ is optimal for the planner. Given the policy (τ_t, ζ_t) in (2.11)-(2.13), consumption c_t^Y is, for $t \geq 1$, given by

$$\begin{aligned} c_t^Y &= \frac{1}{1+\beta} h_t w_t (1-\tau) = \frac{1}{1+\beta}(1+\beta)^{\frac{\theta}{1+\theta}}(1+g)^{t-1} w_1 \frac{1+g}{\lambda w_1}(1+\beta)^{\frac{1}{1+\theta}} \\ &= \frac{(1+g)^t}{\lambda}, \end{aligned}$$

which confirms that (2.3) is satisfied.
Finally, note that if the economy is undergoing a transition, i.e., if the condition $w_1 > (1+g)w_0$ holds, then

$$(1+\beta)^{\frac{1}{1+\theta}} \frac{1+g}{w_1} < \frac{1}{w_0}(1+\beta)^{\frac{1}{1+\theta}},$$

so the range $\lambda \in \left[(1+\beta)^{\frac{1}{1+\theta}} \frac{1+g}{w_1}, \frac{1}{w_0}(1+\beta)^{\frac{1}{1+\theta}}\right]$ is non-empty. □

Proposition B.1. *Consider an economy where wages grow at the constant rate $\tilde{g}$ during the transition and $g < \tilde{g}$ in steady state, i.e., $g_t = \tilde{g}$ for $t \in \{0,1,..,T\}$, and $g_t = g$ for $t > T$. Population growth is assumed to be zero, for simplicty. Agents live for $J \geq 2$ periods and retire after $J_W < J$ periods. The optimal allocation (*first best*) solves the following planning program:*[1]

$$\sum_{t=0}^{\infty} \phi^t \sum_{j=0}^{J} \beta^j \left(\log\left(c_{t,j}\right) - \frac{h_{t,j}^{1+\frac{1}{\theta}}}{1+\frac{1}{\theta}} \right), \tag{B.3}$$

subject to

$$\sum_{t=0}^{\infty} \frac{1}{R^t} \sum_{j=0}^{J} \frac{c_{t,j}}{R^j} = A_0 + \sum_{t=0}^{\infty} \frac{1}{R^t} \sum_{j=0}^{J_w} \frac{w_{t+j} h_{t,j}}{R^j}, \tag{B.4}$$

[1] We ignore for simplicity the generations born before $t = 0$.

where $c_{t,j}$ and $h_{t,j}$ are consumption and labor supply of an individual of age j born at date t. Then, the first-best allocation is given by:

$$c_{t,0} = \lambda^{-1} (\phi R)^t, \tag{B.5}$$

$$c_{t,j} = c_{t,0} (\beta R)^j, \text{ for } j \in \{1,2,...,J\}, \tag{B.6}$$

$$h_{t,j} = \left(\frac{w_{t+j}}{c_{t,j}} \right)^{\theta}, \text{ for } j \in \{0,1,...,J_w\}, \tag{B.7}$$

$$h_{t,j} = 0, \text{ for } j \in \{J_w+1, J_w+2, ..., J\}, \tag{B.8}$$

where λ is a decreasing function of A_0.
Consider a cohort born at k, and let $W_k = \sum_{j=0}^{J_w} \frac{w_{k+j} h_{k,j}}{R^j}$ denote the present value (before-tax) labor income. Denote by $\zeta_k = \sum_{j=J_w+1}^{J} \frac{p_{k,j}}{R^j} / W_k$ as the pension replacement rate, expressed as a share of W_k. The first-best allocation can be implemented by a Ramsey sequence of cohort-specific taxes and pension replacement rates. These sequences have the following characterization:
(1) Taxes are zero in all periods, $\tau_{t,j} = 0$ for all t and j;
(2) The pension replacement sequence satisfy

$$\frac{1+\zeta_{t+1}}{1+\zeta_t} = \left(\frac{\phi R}{1+g} \frac{1+g}{1+\tilde{g}} \right)^{1+\theta} \times F(t), \tag{B.9}$$

where

$$F(t) = \begin{cases} 1 & \text{if} \quad t \leq T - J_w \\ \frac{\sum_{j=0}^{T-t} \beta^j \left(\frac{1+\tilde{g}}{\beta R} \right)^{(1+\theta)\cdot j} + \left(\frac{1+\tilde{g}}{\beta R} \right)^{(1+\theta)\cdot(T-t)} \sum_{j=T-t+1}^{J_w} \beta^j \left(\frac{1+g}{\beta R} \right)^{(1+\theta)\cdot(j-(T-t))}}{\sum_{j=0}^{T-(t+1)} \beta^j \left(\frac{1+\tilde{g}}{\beta R} \right)^{(1+\theta)\cdot j} + \left(\frac{1+\tilde{g}}{\beta R} \right)^{(1+\theta)\cdot(T-(t+1))} \sum_{j=T-t}^{J_w} \beta^j \left(\frac{1+g}{\beta R} \right)^{(1+\theta)\cdot(j-(T-(t+1)))}} & \text{if} \quad t \in \{T-J_w+1, ..., T\} \\ \left(\frac{1+\tilde{g}}{1+g} \right)^{1+\theta} & \text{if} \quad t > T \end{cases} \tag{B.10}$$

is a non-decreasing function of the birth date t. Finally ζ_0 is given by

$$1 + \zeta_0 = \frac{\sum_{j=0}^{J} \beta^j}{\sum_{j=0}^{J_w} \beta^j \left(\frac{w_j}{(\beta R)^j} \right)^{1+\theta}} \times \frac{1}{\lambda^{1+\theta}}. \tag{B.11}$$

Proof. The characterization of the first-best allocation, (B.5)–(B.8) follows from the problem (B.3)-(B.4) using standard methods. Consider, next, the Ramsey policy. Since $\tau_{t,j} = 0$ the agents' intratemporal first-order condition imply equation (B.7). The agents Euler equation imply that $c_{t,j} = (\beta R)^j c_{t,0}$ as in (B.6). Next, plugging in (B.6) and (B.7) into the budget constraint, and recalling that ζ_t is proportional to the present value of earning, yields

$$\sum_{j=0}^{J} \frac{c_{t,0}(\beta R)^j}{R^j} = (1+\zeta_t) \sum_{j=0}^{J_w} \frac{w_{t+j}}{R^j} \left(\frac{w_{t+j}}{(\beta R)^j} \right)^{\theta} c_{t,0}^{-\theta}.$$

Solving for $c_{t,0}$ yields

$$c_{t,0}^{1+\theta} = (1+\zeta_t) \frac{\sum_{j=0}^{J_w} \frac{w_{t+j}}{R^j} \left(\frac{w_{t+j}}{(\beta R)^j} \right)^{\theta}}{\sum_{j=0}^{J} \beta^j}.$$

Lagging the expression, taking the ratio of $c_{t+1,0}/c_t$, and using (B.9)-(B.10), yields

$$\left(\frac{c_{t+1,0}}{c_{t,0}} \right)^{1+\theta} = \left(\frac{\phi R}{1+g} \frac{1+g}{1+\tilde{g}} \right)^{1+\theta} \times F(t) \times \frac{\sum_{j=0}^{J_w} \beta^j \left(\frac{w_{t+1+j}}{(\beta R)^j} \right)^{1+\theta}}{\sum_{j=0}^{J_w} \beta^j \left(\frac{w_{t+j}}{(\beta R)^j} \right)^{1+\theta}}.$$

We now show that replacing $F(t)$ by its expression in (B.10) yields $\frac{c_{t+1,0}}{c_{t,0}} = \phi R$, which is consistent with the optimality condition (B.5). □

Suppose, first, that $t > T$. Then, replacing $F(t)$ by its expression in (B.10) and simplifying terms yields

$$\left(\frac{c_{t+1,0}}{c_{t,0}} \right)^{1+\theta} = \left(\frac{\phi R}{1+g} \frac{1+g}{1+\tilde{g}} \right)^{1+\theta} \times \left(\frac{1+\tilde{g}}{1+g} \right)^{1+\theta} \times (1+g)^{1+\theta} = (\phi R)^{1+\theta},$$

which is consistent with (B.5).

Suppose, next, that $t \in \{T - J_w + 1, ..., T\}$. Then, proceeding as above,

$$\left(\frac{c_{t+1,0}}{c_{t,0}} \right)^{1+\theta} = \left(\frac{\phi R}{1+g} \frac{1+g}{1+\tilde{g}} \right)^{1+\theta} \times$$

$$\frac{\sum_{j=0}^{T-t}\beta^j\left(\frac{1+\tilde{g}}{\beta R}\right)^{(1+\theta)\cdot j}+\left(\frac{1+\tilde{g}}{\beta R}\right)^{(1+\theta)\cdot(T-t)}\sum_{j=T-t+1}^{J_w}\beta^j\left(\frac{1+g}{\beta R}\right)^{(1+\theta)\cdot(j-(T-t))}}{\sum_{j=0}^{T-(t+1)}\beta^j\left(\frac{1+\tilde{g}}{\beta R}\right)^{(1+\theta)\cdot j}+\left(\frac{1+\tilde{g}}{\beta R}\right)^{(1+\theta)\cdot(T-(t+1))}\sum_{j=T-t}^{J_w}\beta^j\left(\frac{1+g}{\beta R}\right)^{(1+\theta)\cdot(j-(T-(t+1)))}}\times\frac{\sum_{j=0}^{J_w}\beta^j\left(\frac{w_{t+1+j}}{(\beta R)^j}\right)}{\sum_{j=0}^{J_w}\beta^j\left(\frac{w_{t+j}}{(\beta R)^j}\right)}$$

Then, simplifying terms yields

$$\left(\frac{c_{t+1,0}}{c_{t,0}}\right)^{1+\theta}=\left(\frac{\phi R}{1+g}\frac{1+g}{1+\tilde{g}}\right)^{1+\theta}\times\left(\frac{w_{t+1}}{w_t}\right)^{1+\theta}=(\phi R)^{1+\theta},$$

which is again consistent with (B.5).

Suppose, finally, that $t\leq T-J_w$. Then, proceeding as above,

$$\left(\frac{c_{t+1,0}}{c_{t,0}}\right)^{1+\theta}=\left(\frac{\phi R}{1+g}\frac{1+g}{1+\tilde{g}}\right)^{1+\theta}\times 1\times(1+\tilde{g})^{1+\theta}=(\phi R)^{1+\theta},$$

which is again consistent with (B.5).

Finally, we show that the individual optimization yields $c_{0,0}=\lambda^{-1}$ proving that the entire Ramsey sequence satisfies the first-best condition (B.5). To this aim, note that

$$c_{0,0}\sum_{j=0}^{J}\beta^j=(1+\zeta_0)\times\sum_{j=0}^{J_w}\beta^j\left(\frac{w_j}{(\beta R)^j}\right)^{1+\theta}c_{0,0}^{-\theta}.$$

Collecting terms and replacing ζ_0 by (B.11) yields $c_{00}=\lambda^{-1}$.

Corollary B.2. *Suppose $\phi=(1+g)/R$. Then, the optimal pension benefit sequence is strictly decreasing for all transition generations, $t\leq T$, and constant for all generations born after the end of the transition, $\zeta_t=\zeta_L$ for all $t>T$.*

$$\frac{1+\tilde{\zeta}_{t+1}}{1+\tilde{\zeta}_t}=\left(\frac{1+g}{1+\tilde{g}}\right)^{1+\theta}.$$

Proof. The proof follows from (B.9)-(B.10), recalling that $\tilde{g}>g$. □

Proposition B.3. *Consider the environment of Proposition B.1. Suppose $\phi=(1+g)/R$, $A_0\geq 0$, and that the Ramsey implementation is subject to the additional constraint that pensions are non-negative, i.e., $\zeta_t\geq 0$ for all t. The second-best Ramsey allocation has the following characterization: Either the constraint $\zeta_t\geq 0$ is never binding (A_0 is very large),*

and the first best can be implemented by the policy descibed in Proposition B.1„ or there exists $\hat{T} < \infty$ *such that:*
(1) If $t < \hat{T}$*, then, up to an increase in* λ *(implying a lower* $c_{0,0}$*), the Ramsey policy sequence is identical to the unconstrained policy sequence that implements first best, i.e., taxes are zero in all periods,* $\tau_{t,j} = 0$ *for all* t *and* j*, and pensions are given by (B.9)—(B.11);*
(2) If $t \geq \hat{T}$*, then,* $\zeta_t = 0$ *and* $\tau_{t,j} = \tau_t > 0$*.*

Proof. The second-best Ramsey problem can be formulated as follows

$$\max_{c,h,\tau,\zeta} \sum_{t=0}^{\infty} \phi^t \sum_{j=0}^{J} \beta^j \left(\log\left(c_{t,j}\right) - \frac{h_{t,j}^{1+\frac{1}{\theta}}}{1+\frac{1}{\theta}} \right), \tag{B.12}$$

subject to:

$$\begin{aligned}
\sum_{t=0}^{\infty} \frac{1}{R^t} \sum_{j=0}^{J} \frac{c_{t,j}}{R^j} &= A_0 + \sum_{t=0}^{\infty} \frac{1}{R^t} \sum_{j=0}^{J_w} \frac{w_{t+j} h_{t,j}}{R^j}, \\
c_{t,j} &= c_{t,0} \left(\beta R\right)^j, \\
h_{t,j} &= \left(1-\tau_{t,j}\right)^{\theta} \left(\frac{w_{t+j}}{\left(\beta R\right)^j} \right)^{\theta} c_{t,0}^{-\theta}, \text{ for } j \leq J_w, \\
h_{t,j} &= 0, \text{ for } j > J_w, \\
\sum \beta^j c_{t,0} &= \left(1+\zeta_t\right) \sum \beta^j \left(1-\tau_{t,j}\right)^{1+\theta} \left(\frac{w_{t+j}}{\left(\beta R\right)^j} \right)^{1+\theta} c_{t,0}^{-\theta}, \\
\zeta_t &\geq 0.
\end{aligned}$$

Using the set of constraints, we can express the second-best problem in terms of the following Lagrangian (note that the Euler equation of consumers allows us to express the problem

as a function of $c_{t,0}$ rather that of the entire consumption sequence of each cohort):

$$L=\sum_{t=0}^{\infty}\phi^{t}\left(\begin{array}{c}\sum_{j=0}^{J}\beta^{j}\log\left(c_{t,0}\left(\beta R\right)^{j}\right)-\sum_{j=0}^{J_w}\beta^{j}\frac{\left(1-\tau_{t,j}\right)^{1+\theta}\left(\frac{w_{t+j}}{(\beta R)^{j}}\right)^{1+\theta}c_{t,0}^{-(1+\theta)}}{1+\frac{1}{\theta}}\\+\mu_{t}\left(\sum_{j=0}^{J}\beta^{j}c_{t,0}-\sum_{j=0}^{J_w}\beta^{j}\left(1-\tau_{t,j}\right)^{1+\theta}\left(\frac{w_{t+j}}{(\beta R)^{j}}\right)^{1+\theta}c_{t,0}^{-\theta}\right)\end{array}\right)+$$
$$\lambda\left(\sum_{t=0}^{\infty}\frac{1}{R^{t}}\sum_{j=0}^{J_w}\beta^{j}\left(1-\tau_{t,j}\right)^{\theta}\left(\frac{w_{t+j}}{(\beta R)^{j}}\right)^{1+\theta}c_{t,0}^{-\theta}-\sum_{t=0}^{\infty}\frac{1}{R^{t}}\sum_{j=0}^{J}\beta^{j}c_{t,0}\right)$$

where $\mu_t \geq 0$ is the Lagrangian multiplier associated with the constraint $\zeta_t \geq 0$, and $\lambda > 0$ is the Lagrange multiplier associated with the resource constraint. □

The FOCs with respect to $c_{t,0}$ and $\tau_{t,j}$ yield, respectively:

$$\frac{\partial L}{\partial c_{t,0}}=\phi^{t}\left(\begin{array}{c}\sum_{j}\beta^{j}\frac{1}{c_{t,0}}+\sum^{J_w}\beta^{j}\theta\left(1-\tau_{t,j}\right)^{1+\theta}\left(\frac{w_{t+j}}{(\beta R)^{j}}\right)^{1+\theta}c_{t,0}^{-(2+\theta)}+\\\mu_{t}\left(\sum\beta^{j}+\theta\sum^{J_w}\beta^{j}\left(1-\tau_{t,j}\right)^{1+\theta}\left(\frac{w_{t+j}}{(\beta R)^{j}}\right)^{1+\theta}c_{t,0}^{-\theta-1}\right)\end{array}\right)-$$
$$\lambda\left(\left(\frac{1}{R^{t}}\sum_{j}^{J_w}\theta\beta^{j}\left(1-\tau_{t,j}\right)^{\theta}\left(\frac{w_{t+j}}{(\beta R)^{j}}\right)^{1+\theta}c_{t,0}^{-\theta-1}+\frac{1}{R^{t}}\sum_{j}\beta^{j}\right)\right)=0, \tag{B.13}$$

$$\frac{\partial L}{\partial \tau_{t,j}}=\phi^{t}\left(\beta^{j}\theta\left(1-\tau_{t,j}\right)^{\theta}\left(\frac{w_{t+j}}{(\beta R)^{j}}\right)^{1+\theta}c_{t,0}^{-(1+\theta)}+\mu_{t}\left(\left(1+\theta\right)\beta^{j}\left(1-\tau_{t,j}\right)^{\theta}\left(\frac{w_{t+j}}{(\beta R)^{j}}\right)^{1+\theta}c_{t,0}^{-\theta}\right)\right)-$$
$$\lambda\left(\frac{1}{R^{t}}\theta\beta^{j}\left(1-\tau_{t,j}\right)^{\theta-1}\left(\frac{w_{t+j}}{(\beta R)^{j}}\right)^{1+\theta}c_{t,0}^{-\theta}\right)=0. \tag{B.14}$$

Consider, next, two separate cases:

1. $\mu_t = 0$, i.e., the constraint $\zeta_t \geq 0$ is slack. In this case, the problem is identical to the implementation of the first best in Proposition B.1, up to an increase in the value of λ. In particular, letting $\tau_{t,j} = \tau_t = 0$ implies that $c_{t,0} = \lambda^{-1}(\phi R)^{t}$ (cfr. equation (B.5)) and $h_{t,j} = \left(\frac{w_{t+j}}{c_{t,j}}\right)^{\theta}$, for $j \in \{0,1,...,J_w\}$ (cfr. equation (B.7)). Since λ is larger, consumption is lower and labor supply is higher. Moreover, if the constraint is slack at $t > 0$, it must also be slack for all $k \leq t$. To see why, note that the pension

sequence ζ_t given by (B.9)-(B.11) is non-increasing, so $\zeta_t > 0$ (and, thus, $\mu_t = 0$) implies $\zeta_k > 0$ (thus, again, $\mu_k = 0$) for all $k < t$.

2. $\mu_t > 0$, i.e., the constraint that pensions cannot be negative is binding. Thus, $\zeta_t = 0$ and the individual budget constraint yields:

$$\sum \beta^j c_{t,0} = \sum_{j}^{J_w} \beta^j \left(1-\tau_{t,j}\right)^{1+\theta} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} c_{t,0}^{-\theta} \tag{B.15}$$

Combining (B.13)-(B.14) yields:

$$\phi^t \left(\sum_j \beta^j \frac{1}{c_{t,0}} + \mu_t \left(\sum \beta^j - \sum_{j}^{J_w} \beta^j \left(1-\tau_{t,j}\right)^{1+\theta} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} c_{t,0}^{-\theta-1} \right) \right) - \lambda \left(\frac{1}{R^t} \sum_j \beta^j \right) = 0.$$

Substituting into this expression the budget constraint, (B.15), implies:

$$\phi^t \sum_j \beta^j \frac{1}{c_{t,0}} - \lambda \frac{1}{R^t} \sum_j \beta^j = 0 \Rightarrow$$

$$c_{t,0} = \lambda^{-1} \left(\phi R\right)^t.$$

Finally, substituting this condition into (B.14), and solving for τ_t, after rearranging terms, yields:

$$\tau_{t,j} = \tau_t = \frac{\mu_t \left(1+\theta\right) c_{t,0}}{\theta + \mu_t \left(1+\theta\right) c_{t,0}} > 0,$$

where the inequality follows from the assumption that $\mu_t > 0$. Finally, we can prove by *reductio ad absurdum* that if $\mu_t > 0$, then $\mu_k > 0$ for all $k > t$. Suppose not, and $\exists k > t$ such that $\zeta_k > 0$. Then, for the argument provided in the proof of part 1 of this proposition, the non-negativity constraint should be slack for all $k' < k$, including $k' = t$, raising a contradiction.

3. Finally, note that either the constraint $\zeta_t \geq 0$ is slack for all T, and then the first best can be implemented, or there exist T such that the constraint is slack for all $t < T$ and is binding for all $t \geq T$.

B.2 Estimation method of the rural-urban migration

In this appendix, we present the estimation method of the rural-urban migration. $n_{2000}^{h,i,j}$ and $n_{2005}^{h,i,j}$ represent the population of group (h,i,j) in the 2000 census and 2005 survey, respectively, where $h \in \{u,r\}$, $i \in \{f,m\}$, and $j \in \{0,1,\cdots,100\}$ stand for residential status (u for urban and r for rural residents), gender (f for females and m for males), and age, respectively. $\hat{n}_{2005}^{h,i,j}$ represents the projected "natural" population in 2005. Denote $m^{i,j}$ the net flow of the rural-urban migration from 2000 to 2005. We observe $n_{2000}^{h,i,j}$ and $n_{2005}^{h,i,j}$ from the 2000 census and 2005 survey. Moreover, we can use $n_{2000}^{h,i,j}$, together with the observed birth and mortality rates, to project $\hat{n}_{2005}^{h,i,j}$; i.e., the "natural" population in 2005. In other words, both $n_{2005}^{h,i,j}$ and $\hat{n}_{2005}^{h,i,j}$ in (B.16) and (B.17) are observable. The 2005 urban and rural population gender-age structure can thus be composed into three parts:

$$n_{2005}^{u,i,j} = \hat{n}_{2005}^{u,i,j} + m^{i,j} + \varepsilon^{u,i,j}, \tag{B.16}$$

$$n_{2005}^{r,i,j} = \hat{n}_{2005}^{r,i,j} - m^{i,j} + \varepsilon^{r,i,j}, \tag{B.17}$$

where $\varepsilon^{h,i,j}$ captures measurement errors in the census and survey.

In the ideal case with no measurement errors, either (B.16) or (B.17) can back out $m^{i,j}$. The measurement error on the total population, $\sum_{h,i,j}\varepsilon^{h,i,j}$, is small. When $\sum_{h,i,j}\varepsilon^{h,i,j}=0$, (B.16) and (B.17) imply that the projected total population, $\sum_{h,i,j}\hat{n}_{2005}^{h,i,j}$, would be equal to the total population in the 2005 survey, $\sum_{h,i,j}n_{2005}^{h,i,j}$. The difference between $\sum_{h,i,j}\hat{n}_{2005}^{h,i,j}$ and $\sum_{h,i,j}n_{2005}^{h,i,j}$ is less than 1%. [2] However, the match of the sum of the rural and urban population in each gender-age group is less perfect. Figure A-1 plots the projected 2005 "natural" population gender-age structure (solid line) and the 2005 survey data (dotted line). The discrepancy between the two lines reveals the measurement error on the population of each gender-age group, $\varepsilon^{i,j}$, where

$$\varepsilon^{i,j} \equiv \sum_h \varepsilon^{h,i,j} = \sum_h \left(n_{2005}^{h,i,j} - \hat{n}_{2005}^{h,i,j} \right). \tag{B.18}$$

[2]Despite the small discrepancy, to avoid biased estimates, we adjust $n_{2000}^{h,i,j}$ by a scale of κ, where κ is calibrated to 1.0073 by matching the projected 2005 total population with the 2005 survey data. $\kappa = 1.0073$ suggests the discrepancy of the total population to be less than 1%.

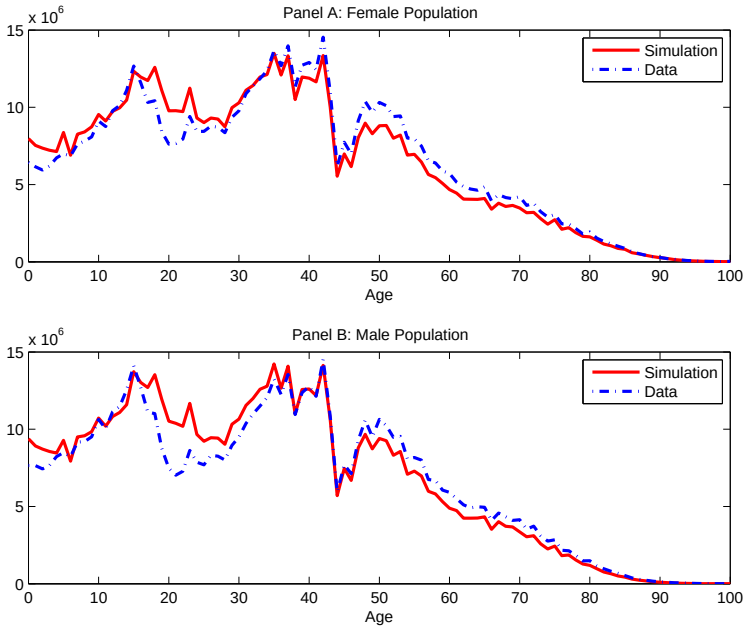

Figure B.1: Population in 2005
The upper panel shows the female population of different ages in 2005, in the survey data (solid line), and in our simulation (dashed line). The lower panel shows the male population in 2005.

Figure A-1 suggests $\varepsilon^{i,j}$ to be quantitatively important.[3] To understand how $\varepsilon^{i,j}$ affects the estimated migration gender-age structure, let us assume the measurement error on urban population, $\varepsilon^{u,i,j}$, is proportional to $\varepsilon^{i,j}$:

$$\varepsilon^{u,i,j} = \pi \cdot \varepsilon^{i,j}, \tag{B.19}$$

where $\pi \in [0,1]$. It follows that the measurement error on rural population is

$$\varepsilon^{r,i,j} = (1-\pi) \cdot \varepsilon^{i,j}. \tag{B.20}$$

[3]If all the discrepancies are due to sampling errors in the 2005 survey, the comparison between the two lines in Figure A-1 indicates that a major drawback of the 2005 survey is the undercounted young labor force (age 16 to 40). Our calculation suggests 66 million young labor force (11% of total young labor force) missing from the 2005 survey.

Rearranging (B.16) gives the net flow of migration:

$$\sum_i \sum_j m^{i,j} = \sum_i \sum_j \left(n^{u,i,j}_{2005} - \hat{n}^{u,i,j}_{2005} \right) - \pi \sum_i \sum_j \varepsilon^{i,j} \tag{B.21}$$

$$= \sum_i \sum_j \left(n^{u,i,j}_{2005} - \hat{n}^{u,i,j}_{2005} \right) - \pi \sum_h \sum_i \sum_j \left(n^{h,i,j}_{2005} - \hat{n}^{h,i,j}_{2005} \right). \tag{B.22}$$

The second equality comes from (B.18). Let us consider two extreme cases of π. When $\pi = 1$, (B.21) can be written as

$$\sum_i \sum_j m^{i,j} = \underbrace{\sum_i \sum_j \hat{n}^{r,i,j}_{2005}}_{\text{projected "natural" rural population}} - \underbrace{\sum_i \sum_j n^{r,i,j}_{2005}}_{\text{rural population in the survey data}}.$$

When $\pi = 0$, (B.21) reduces to

$$\sum_i \sum_j m^{i,j} = \underbrace{\sum_i \sum_j n^{u,i,j}_{2005}}_{\text{urban population in the survey data}} - \underbrace{\sum_i \sum_j \hat{n}^{u,i,j}_{2005}}_{\text{projected "natural" urban population}}.$$

Therefore, the choice of π boils down to the choice of using rural or urban population to back out migration. It has been widely acknowledged that urban population survey tends to underestimate "floating population," that is, rural migrants without *hukou* - the local household registration status (e.g., Liang and Ma, 2004). So, we set $\pi = 1$. We will discuss the results using $\pi = 0.5$.

It is instructive to compare the actual migration structure with our estimates. The migration *flow* structure is hard to obtain. However, the migration *stock* structure may shed some light on the *flow* structure. The age structure of migrants in the 2000 census is presented in the second row of Table A-1, which has a high concentration in the 15-29 age group. The same pattern also appears in our estimates under $\pi = 1$ (the third row). $\pi = 0.5$ results in a much more dispersed age structure (the fourth row). This provides a justification for using $\pi = 1$.[4]

[4]One caveat is that the data from the 2000 census are the age structure of narrowly defined migrants, whereas our estimate is on broadly defined migrants including urbanized population.

Table A-1 Age distribution of migration (percent)

age	<15	15-29	30-44	45-59	60+
migration stock in the 2000 census	9.0	60.5	22.2	5.8	2.5
estimated flow from 2000 to 2005 with $\pi = 1$	25.8	64.8	26.5	-8.6	-8.6
estimated flow from 2000 to 2005 with $\pi = 0.5$	17.8	39.5	27.7	8.9	6.1

Note: The age structure in the 2000 census is from (Liang and Ma, 2004).

Finally, we compute $mr^{i,j}$, the age–gender specific migration rate defined as the average annual net flow of migration per hundred rural population with gender i and age j. We assume that $mr^{i,j}$ is time-invariant and the mortality rates for migrants are the same as those for rural residents. Then, $m^{i,j}$ can be written as follows:

$$
\begin{aligned}
m^{i,j} \;=\; & \underbrace{mr^{i,j-5} n_{2000}^{r,i,j-5}}_{\text{migration of 2000}} \underbrace{\left(1-d_{2000}^{r,i,j-1}\right)\cdots\left(1-d_{2000}^{r,i,j-5}\right)}_{\text{survival rate from 2000 to 2005}} \\
& + \underbrace{mr^{i,j-4}\left(1-mr^{r,j-5}\right) n_{2000}^{r,i,j-5}}_{\text{migration of 2001}} \underbrace{\left(1-d_{2000}^{r,i,j-1}\right)\cdots\left(1-d_{2000}^{r,i,j-5}\right)}_{\text{survival rate from 2001 to 2005}} \\
& + \underbrace{mr^{i,j-3}\left(1-mr^{r,j-4}\right)\left(1-mr^{r,j-5}\right) n_{2000}^{r,i,j-5}}_{\text{migration of 2002}} \underbrace{\left(1-d_{2000}^{r,i,j-1}\right)\cdots\left(1-d_{2000}^{r,i,j-5}\right)}_{\text{survival rate from 2001 to 2005}} \\
& + \underbrace{mr^{i,j-2}\left(1-mr^{r,j-3}\right)\left(1-mr^{r,j-4}\right)\left(1-mr^{r,j-5}\right) n_{2000}^{r,i,j-5}}_{\text{migration of 2003}} \underbrace{\left(1-d_{2000}^{r,i,j-1}\right)\cdots\left(1-d_{2000}^{r,i,j-5}\right)}_{\text{survival rate from 2001 to 2005}} \\
& + \underbrace{mr^{i,j-1}\left(1-mr^{r,j-2}\right)\cdots\left(1-mr^{r,j-5}\right) n_{2000}^{r,i,j-5}}_{\text{migration of 2004}} \underbrace{\left(1-d_{2000}^{r,i,j-1}\right)\cdots\left(1-d_{2000}^{r,i,j-5}\right)}_{\text{survival rate from 2001 to 2005}}.
\end{aligned}
$$

Here, $n_{2000}^{r,i,j-5}$ is the mortality rate of rural residents in the 2000 census. In other words, $m^{i,j}$ measures an accumulated migration stock from 2000 to 2005. The above equation allows

us to back out the age-gender specific migration rates. Specifically, for $j = J+5$:

$$m^{i,J+5} = \underbrace{mr^{i,J}\hat{n}^{r,i,J}_{2000}}_{\text{migration of 2000}} \underbrace{\left(1-d^{r,i,J+4}_{2000}\right)\cdots\left(1-d^{r,i,J+4}_{2000}\right)}_{\text{survival rate from 2000 to 2005}}$$

$\Rightarrow$

$$mr^{i,J} = \frac{m^{i,J+5}}{n^{r,i,J}_{2000}\left(1-d^{r,i,J+4}_{2000}\right)\cdots\left(1-d^{r,i,J}_{2000}\right)}.$$

For $j = J+4$:

$$\begin{aligned} m^{i,J+4} &= \underbrace{mr^{i,J-1}\hat{n}^{r,i,J-1}_{2000}}_{\text{migration of 2000}}\underbrace{\left(1-d^{r,i,J+3}_{2000}\right)\cdots\left(1-d^{r,i,J-1}_{2000}\right)}_{\text{survival rate from 2000 to 2005}} \\ &\quad + \underbrace{mr^{i,J}\left(1-mr^{r,J-1}\right)n^{r,i,J-1}_{2000}}_{\text{migration of 2001}}\underbrace{\left(1-d^{r,i,J+3}_{2000}\right)\cdots\left(1-d^{r,i,J-1}_{2000}\right)}_{\text{survival rate from 2000 to 2005}} \end{aligned}$$

$\Rightarrow$

$$mr^{i,J-1} = \frac{m^{i,J+4} - mr^{i,J}n^{r,i,J-1}_{2000}\left(1-d^{r,i,J+3}_{2000}\right)\cdots\left(1-d^{r,i,J-1}_{2000}\right)}{\left(1-mr^{i,J}\right)n^{r,i,J-1}_{2000}\left(1-d^{r,i,J+3}_{2000}\right)\cdots\left(1-d^{r,i,J-1}_{2000}\right)}.$$

All the migration rates can thus be solved in a recursive way.

B.3 Details on the Chinese pension system

This appendix provides a description of the basic features of the Chinese pension system. We start with the urban pension system, and then provide a brief description of the rural pension system, which has been introduced experimentally in 2009.

B.3.1 The urban pension system

The pre-1997 urban pension system was primarily based on state and urban collective enterprises in a centrally planned economy. Retirees received pensions from their employers, with replacement rates that could be as high as 80 percent (see, e.g., Sin, 2005; OECD,

2007). The coverage was low in the work-unit-based system, though. Many non-state-owned enterprises had no pension scheme for their employees. The coverage rate, measured by the ratio of the number of workers covered by the system to the urban employment, was merely 44% in 1992 according to *China Statistical Yearbook* 2009. The rapid expansion of the private sector caused a growing disproportion between the numbers of contributors and beneficiaries and, therefore, a severe financial distress for the old system (Zhao and Xu, 2002). To deal with the issue, the government initiated a transition from the traditional system to a public pension system in the early 1990s. The new system was implemented nationwide after the State Council issued "A Decision on Establishing a Unified Basic Pension System for Enterprise Workers (Document 26)" in 1997.

The reformed system mainly consists of two pillars. The first pillar, funded by 17% wage taxes paid by enterprises, guarantees a replacement rate of 20% of local average wage for retirees with a minimum of 15 years of contribution. It is worth emphasizing that the pension fund is managed by local governments (previously at the city level and now at the provincial level). The second pillar provides pensions from individual accounts financed by a contribution of 3% and 8% wage taxes paid by enterprises and workers, respectively. There is a third pillar adding to individual accounts through voluntary contribution. The return of individual accounts is adjusted according to bank deposit rates. The system also defines monthly pension benefits from individual accounts equaling the account balance at retirement divided by 120. The targeted replacement rate of the system is 58.5%.[5]

More recently, a new reform was implemented after the State Council issued "A Decision on Improving the Basic Pension System for Enterprise Workers (Document 38)" in 2005. The reform adjusted the proportion of taxes paid by enterprises and individuals and the proportion of contribution for individual accounts. Individual accounts are now funded by the wage taxes of 8% paid by workers only.[6]

[5]Suppose that the wage growth rate is equal to the interest rate. For a worker who contributes to the system for 35 years (from age 25 to 60), her pension benefits should be equal to 20% of the local average wages (the first pillar) plus 38.5% of her wage before retirement.

[6]The reform also adjusted the pension benefits. The replacement rate of an individual is now determined by years of contribution: A one year contribution increases the replacement rate of a wage index averaged from local and individual wages by one percentage point. However, the article did not state explicitly how to compute the wage index.

In practice, the index appears to differ across provinces. For instance, the increase in the average pension benefits per retiree in 2011 was almost the same across Beijing and GanSu (the monthly increase was

Two features of the current urban pension system is particularly important for our modeling. First, the pension reform was cohort-specific. There were three types of cohorts when the pension reform took place: Cohorts enter into the labor market after 1997 (*Xinren*), cohorts retired before 1997 (*Laoren*) and cohorts in between (*Zhongren*). Pension contributions and benefits of *Xinren* are entirely determined by the new rule. According to Item 5 in Document 26, the government commits to pay *Laoren* the same pension benefits as those in the old system subject to an annual adjustment by wage growth and inflation. For *Zhongren*, their contributions follow the new rule, while their benefits consist of two components: (1) pensions from the new system identical to those for *Xinren,* and (2) a transitional pension that smooths the pension gap between *Laoren* and *Xinren*. For simplicity, we ignore *Zhongren* and take pensioners retiring before and after 1997 as *Laoren* and *Xinren*, respectively. Following Sin (2005), we set the replacement rate for *Laoren* and *Xinren* to 78% and 60%, respectively.

Second, like private savings, pension funds are allowed to invest in domestic stock markets. The baseline model assumes the annual rate of returns to pension funds to be 2.5%, which is identical to the rate of returns to private savings. According to the latest information released by the National Council for Social Security Fund, the average share of pension funds invested in stock markets was 19.22% in 2003-2011.[7] If 20% of pension funds have access to the market with an annual return of 6% and the rest of the funds gain an annual return of 1.75% as the one-year bank deposits, the average annual rate of returns would be equal to 2.6%, almost equal to 2.5% set in the baseline model.

It is also worth emphasizing that the actual urban pension system deviates from statutory regulations in a number of ways and our model has been adapted to capture some major discrepancies. First, the individual accounts are basically empty. Despite the recent efforts made by the central government to fund these empty individual accounts, there are only 270 billion RMB in all individual accounts of around 200 million workers participating in the urban pension system.[8] Therefore, we take the individual accounts as notional

RMB210 in Beijing and RMB196 in GanSu), though the average wage in Beijing is more than two times as high as that in GanSu and the gap has been rather stable over time.

[7] Source: http://www.ssf.gov.cn/xw/xw_gl/201205/t20120509_4619.html.

[8] The number of 270 billion RMB comes from the information released by the Ministry of Human Resources and Social Security in the 2012 National People's Congress. Source: http://lianghui.people.com.cn/2012npc/GB/239293/17320248.html

and ignore any distinction between the different pension pillars throughout the paper. In addition, we assume that 40% of pension benefits are indexed to wage growth. The level of indexation is set on the conservative side since the actual level is between 40% and 60% (see Sin, 2005).

Second, the statutory contribution rate including both basic pensions and individual accounts is 28%, of which 20% should be paid by firms and 8% should be paid by workers (see the above discussion on Document 26 and 38). However, there is evidence that a significant share of the contributions is evaded. For instance, in the annual National Industrial Survey – which includes all state-owned manufacturing enterprises and all private manufacturing enterprises with revenue above 5 million RMB – the average pension contributions paid by firms in 2004-2007 amounts to 11% of the average wages, 9 percentage points below the statutory rate.[9] Most evasion comes from privately owned firms, whose contribution rate is a merely 7%.

The actual contribution rate is substantially lower than the statutory rate even for workers participating in the system. A simple way of estimating the actual contribution rate conditional on participation is to look at the following ratio:

$$\begin{aligned} BR &\equiv \frac{\text{per retiree pension benefits}}{\text{per worker pension contributions}} \\ &\equiv \frac{\frac{\text{total pension fund expenditure}}{\text{total retirees covered by the system}}}{\frac{\text{total pension fund revenue - government subsidy}}{\text{total workers covered by the system}}}. \end{aligned}$$

If the replacement rate is indeed 60%, a contribution rate of 28% would imply *BR* to be 2.1. However, we find that the average *BR* in the data from 1997 to 2009 is 3.1, much higher than 2.1 by the statutory contribution rate. With a targeted replacement rate of 60%,

[9] In addition, with a labor income share less than 20%, wages appear to be severely underreported.

the ratio of 3.1 would imply an actual contribution rate of 19.4%.[10] So, we set the actual contribute rate to 20% in the paper.

Finally, although the coverage rate of the urban pension system is still relatively low, it has grown from about 40% in 1998 to 57% in 2009, where we measure the coverage rate by the number of employees participating in the pension system as a share of the number of urban employees.[11] There is a concern that the rapidly growing size of migrant workers might lead to downward-biased urban employment. Our estimation suggests that the urban population (including migrants) between age 22 and 60 increases by 130 million from 2000 to 2009. A labor participation rate of 80% would imply an increase of 104 million in the urban employment, whereas the increase by the official statistics is 79 million. Restoring the 25 million "missing" urban employment would lower the pension coverage rate from 57% to 53% in 2009. Our baseline model assumes a constant coverage rate of 60%, reflecting a trade-off between the low coverage of the current pension system and the potentially higher one in the future.

B.3.2 The rural pension system

The pre-2009 rural pension program had two features. First, it was "fully-funded" in the sense that pension benefits were essentially determined by contributions to individual accounts. Second, the coverage rate was low since farmers did not have incentives to participate. A pilot pension program was launched for rural residents in 2009. Like those in the urban pension system, the new rural program entails two benefit components. The first one is referred to as basic pension, mainly financed by the Ministry of Finance, and the second one is pension from individual account. If a migrant worker who joined the urban pension system returns to her home town, the money accumulated in her account will be transferred to her new account in the rural pension program. The program was first implemented in

[10] All the data are available from *China Statistical Yearbook*, except for the government subsidies. Fortunately, since 2010, the Ministry of Finance has started to publicize detailed expenditure items. The government subsidy to the pension fund amounted to 191 billion RMB in 2010, accounting for 21% of the total government social security and employment expenditure. We then use 21% to back out annual government subsidy to pension funds from annual total government social security and employment expenditure, which is available from *China Statistical Yearbook*.

[11] Both numbers are obtained from *China Statistical Yearbook* 2010.

10% of cities and counties on a trial basis. The government targeted to extend the program to 60% of cities and counties in 2011. Many of the cities and counties report high participation rates (above 80%). This is not surprising since the program is heavily subsidized (see below for more details).

We then lay out some basic features of the new program upon which the model is based. According to "Instructions on New Rural Pension Experiments" issued by the State Council in 2009, the new program pays a basic pension of RMB55 ($8.7) per month. Suppose that the rural wage equals the rural per capita annual net income, which was RMB5153 in 2009 (*China Statistical Yearbook* 2010). Then, the basic pension would correspond to a replacement rate of 12.8%. Notice that provinces are allowed to choose more generous rural pensions. So, the replacement rate of 9% should be viewed as a lower bound.[12] In practice, some places set a much higher basic pension standard. Beijing, for instance, increased the level to RMB280. The monthly basic pension in Shanghai has a range from RMB150 to RMB300, dependent of age, years of contribution and status in the old pension program.[13] Since the rural per capita net income in Beijing and Shanghai is about 1.4 times higher than the average level in China, a monthly pension of RMB280 would imply a replacement rate of 27.2%. In the quantitative exercise, we then set the replacement rate to 20% to match the average of the basic level of 12.8% and the high level of 27.2%.[14] On the contribution side, rural residents in principle should contribute 4% to 8% of the local average income per capita in the previous year. We take the mean and set a contribution rate of 6%.[15]

[12]The Ministry of Human Resources and Social Security has made it clear that there is no upper bound for basic pension and local governments may increase basic pension according to their public financing capacity.

[13]See "Detailed Rules for the Implementation of Beijing Urban-Rural Household Pension Plans," Beijing Municipal Labor and Social Security Bureau, 2009 and "Implementation Guidelines of State Council's Instructions on New Rural Pension Experiments," Shanghai Municipal Government, 2010.

[14]All rural residents above age 60 are entitled to basic pension. The only condition is that children of a basic pension recipient, if any, should participate in the program. In practice, basic pension might be contingent on years of contribution and status in the old pension program (see the above example from Shanghai).

In addition, a recent official policy report from the Ministry of Human Resources and Social Security (http://news.qq.com/a/20090806/000974.htm) states that by the rule of the new system, a rural worker paying an annual contribution rate of 4% for 15 years should be entitled to pension benefits with a replacement rate of 25%.

[15]Rural residents are allowed to contribute more. But the contribution rate cannot exceed 15% for each person. Moreover, to be eligible for pension from individual account, a rural resident must contribute to

The current pension program heavily relies on government subsidy. *China Statistical Yearbook* 2010 reports a rural population of 712.88 million. According to the 2005 one-percent population survey, 13.7% of rural population is above age 60. These two numbers give a rural population of 97.66 million who are entitled to basic pension. This, in turn, implies an annual government subsidy of 64.46 billion RMB, if monthly basic pension is set to RMB55. The central government revenue is 3592 billion RMB in 2009. So, a full-coverage rural pension program in 2009 would require subsidy as a share of the central government revenue of 1.8% and a share of GDP of 0.19%.

the program for at least 15 years. The monthly pension benefit is set equal to the accumulated money in individual account divided by 139 (the same rule applied to the urban pension program).

Appendix C

Appendix of Chapter 3

In the appendix, we provide proofs for the propositions and some lemmata that are useful.

First, we prove the propositions in the static model. The results in proposition 3.2 are already established in the main text. Here we provide the proof for proposition 3.3, which states the equilibrium outcomes conditional on sufficiently large revolution cost.

Proof. If the cost of revolution for incumbent citizens is high, they only prefer revolution to peace when the trust is very high, potentially higher than the cut-off value of war and peace. This is to say,

$$\begin{aligned} p^R &> p^W \iff \\ \frac{\pi^P\theta\tau y_O - \left(1-\pi^P\theta\right)\tau y_I + f^R}{1-\pi^P\theta} &> \frac{\left(\pi^W-\pi^P\right)\theta\tau(y_I+y_O) - \pi^W\theta\tau f^W}{1-\tau+2\pi^P\theta\tau} \iff \\ f^R &> \frac{\left(1-\pi^P\theta\right)\left(\left(\pi^W-\pi^P\right)\theta\tau(y_I+y_O) - \pi^W\theta\tau f^W\right)}{1-\tau+2\pi^P\theta\tau} \\ & \quad -\pi^P\theta\tau y_O + \left(1-\pi^P\theta\right)\tau y_I \\ &\doteq \underline{f^R}. \end{aligned}$$

If the cost of revolution on the elite is high enough, elite doesn't want revolution at all, since it is always dominated by either war or peace. Let us consider the case that revolution

is always dominated by war, as the following:

$$\begin{aligned} y_E^R &< y_E^W \iff \\ -f_E^R &< \pi^W(1-\theta)\tau(y_I+y_O-f^W) \iff \\ f_E^R &> -\pi^W(1-\theta)\tau(y_I+y_O-f^W) \\ &\doteq \underline{f_E^R}. \end{aligned}$$

□

Lemma C.1. *The value function of the elite's lifetime income $V(p)$ is bounded above and below by some $\bar{V}$ and $\underline{V}$, respectively.*

Proof. To find the solution for the value function, we first characterize important properties of the value function. First, it is bounded. If the expected trade surplus is very close to its lower bound 0, the elite can always at least start a war and gain from the tax extraction, so we get a lower bound of V:

$$\begin{aligned} V &\geq y_E^W + \beta\pi^W y_E^W + \beta^2\pi^W y_E^W + ... \\ &= \frac{1}{1-\beta\pi^W} y_E^W \\ &\doteq \underline{V}. \end{aligned}$$

Similarly, if in each period, the elite expects to get at most $\pi^W(1-\theta)\tau(y_I+y_O+2) \doteq \bar{y}_E$ when the expected trade surplus is as high as 1 for each group, and the probability of the elite staying in power is at most π^W, so we can get a upper bound of V as follows:

$$\begin{aligned} V &\leq \bar{y}_E + \beta\pi^W \bar{y}_E + \beta^2\pi^W \bar{y}_E + ... \\ &= \frac{1}{1-\beta\pi^W} \bar{y}_E \\ &\doteq \bar{V}. \end{aligned}$$

□

With this lemma, we can first prove and discuss proposition 3.5 which shows that war occurs when the trust is low.

Proof. War dominates peace if and only if:

$$\begin{aligned}
V^P(p) &< V^W(p) \iff \\
&\quad \pi^P(1-\theta)\tau(y_I+y_O+2p)+\beta\pi^P E^P\left[V\left(p'\right)\right] \\
&< \pi^W(1-\theta)\tau(y_I+y_O-f^W)+\beta\pi^W E^W\left[V\left(p'\right)\right] \iff \\
p &< \frac{\left(\pi^W-\pi^P\right)(1-\theta)\tau(y_I+y_O)+\beta\pi^W E^W\left[V\left(p'\right)\right]-\beta\pi^P E^P\left[V\left(p'\right)\right]-\pi^W(1-\theta)\tau f^W}{2\pi^P(1-\theta)\tau} \\
&\doteq \underline{p}(p).
\end{aligned}$$

This means that if p is sufficiently small, the gain from trade is smaller than the benefit of war $\underline{p}(p)$, which contains three parts: (1) higher probability of staying in power and getting the current period income: $\left(\pi^W-\pi^P\right)(1-\theta)\tau(y_I+y_O)$; (2) difference in continuation value in the future between war and peace: $\beta\pi^W E^W\left[V\left(p'\right)\right]-\beta\pi^P E^P\left[V\left(p'\right)\right]$; and (3) the cost of war $\pi^W(1-\theta)\tau f^W$. If the cost the war f^W is small, $\underline{p}(p)$ is large, and if meanwhile the trust is low, we have $p<\underline{p}(p)$, implying that the elite chooses war instead of peace. Consider the case that $f^W<\frac{\left(\pi^W-\pi^P\right)}{\pi^W}(y_I+y_O)\doteq\bar{f^W}$. If the trust is at the minimal level, i.e., $p=0$, we have $\underline{p}(0)>p=0$, and the elite prefers war. We can verify this in the following three steps. First, $E^W\left[V\left(p'\right)\right]=E^P\left[V\left(p'\right)\right]=V(\phi)$, as $p=0$ implies $p^+=0$ in both war and peace, and then $p'=\phi$ due to the possibility of type change. Second, $V(\phi)>0$. We know that $\phi\leq\frac{\phi}{\phi+\psi}\leq p^R$, and given $p=\phi$, one possible choice for the elite is to keep having war forever and there will be no threat of revolution, as p will converges to $\frac{\phi}{\phi+\psi}\leq p^R$. In his case, every period, the elite's expected income is $y_E^W=\pi^W(1-\theta)\tau(y_I+y_O-f^W)>0$. The optimal choice for the elite gives higher life-time income than permanent war, so $V(\phi)>0$. This is in fact true for all $p\leq p^R$. Third,

$$\begin{aligned}
p(0) &= \frac{\left(\pi^W-\pi^P\right)(1-\theta)\tau(y_I+y_O)-\pi^W(1-\theta)\tau f^W}{2\pi^P(1-\theta)\tau}+\frac{\beta\pi^W\left[V(\phi)\right]-\beta\pi^P\left[V(\phi)\right]}{2\pi^P(1-\theta)\tau} \\
&> 0+0 \\
&= 0.
\end{aligned}$$

This shows that when the trust is as low as 0, the elite prefers war. Moreover, if $V(p)$ is continuous at $p=\phi$, $\underline{p}(p)$ is continuos at $p=0$. Then given $0<\underline{p}(0)$, there exists a

neighborhood of 0, denoted as $[0, p_E^W)$, for any $p \in [0, p_E^W)$, we have $p < \underline{p}(p)$. In other words, in this low trust region, war is started because there is little to gain from trade. □

Proposition 3.6 shows that "surprisingly", the war also occurs when the trust is too high, because the elite is afraid of even higher trust leading to revolution. We provide the proof and the discussion below.

Proof. Consider the situation $p = p^R$. If peace is chosen, with probability $q^R = p^R q_H + \left(1 - p^R\right)(1 - q_L)$, the trade outcome is good and $p' > p^R$, which triggers a revolution in the next period. Then

$$V^P\left(p^R\right) \leq \pi^P(1-\theta)\tau(y_I + y_O + 2p^R) + \beta\pi^P\left(-q^R f_E^R + \left(1-q^R\right)\bar{V}\right).$$

A sufficient condition for $V^P\left(p^R\right) < V^W\left(p^R\right)$ is:

$$\begin{aligned}
& \pi^P(1-\theta)\tau(y_I + y_O + 2p^R) + \beta\pi^P\left(-q^R f_E^R + \left(1-q^R\right)\bar{V}\right) \\
< \;& \pi^W(1-\theta)\tau(y_I + y_O - f^W) + \beta\pi^W\underline{V} \iff \\
f_E^R \;>\; & \frac{\pi^P(1-\theta)\tau(y_I + y_O + 2p^R) - \pi^W(1-\theta)\tau(y_I + y_O - f^W) + \beta\pi^P\left(1-q^R\right)\bar{V} - \beta\pi^W\underline{V}}{\beta\pi^P q^R} \\
\doteq \;& \underline{f_E^R}.
\end{aligned}$$

We can see that given f_E^R is sufficiently large, the elite chooses to go to war when the trust is as high as p^R. In fact, we can see that if p increases from the below, before it reaches p^R, war is already preferred by the elite to peace. This is because if the trust is slightly smaller than p^R, the threat of revolution in peace - the probability of having a high enough trust that leads to revolution - is only slightly smaller and the cost of peace for the elite is still large. In other words, there exists a neighborhood of p^R, denoted as $(p_E^R, p^R]$, for all p in this neighborhood, we have

$$\begin{aligned}
f_E^R \;>\; & \frac{\pi^P(1-\theta)\tau(y_I + y_O + 2p) - \pi^W(1-\theta)\tau(y_I + y_O - f^W) + \beta\pi^P\left(1-q^R\right)\bar{V} - \beta\pi^W\underline{V}}{\beta\pi^P q^R} \Longrightarrow \\
V^P\left(p^R\right) \;<\; & V^W\left(p^R\right).
\end{aligned}$$

War is preferred by the elite when $p > p_E^R$. □

Bibliography

Abel, A. B., Mankiw, N. G., Summers, L. H., and Zeckhauser, R. J. (1989). Assessing dynamic efficiency: Theory and evidence. *Review of Economic Studies*, 56(1):1–19.

Acemoglu, D. (2003). Why Not a Political Coase Theorem? Social Conflict, Commitment, and Politics. *Journal of Comparative Economics*, 31(4):620–652.

Acemoglu, D. (2008). Oligarchy Versus Democracy. *Journal of European Economic Association*, 6(March):1–44.

Acemoglu, D., Aghion, P., and Zilibotti, F. (2006a). Distance to Frontier, Selection, and Economic Growth. *Journal of European Economic Association*, 4(March):37–74.

Acemoglu, D., Aghion, P., and Zilibotti, F. (2006b). Distance to frontier, selection, and economic growth. *Journal of the European Economic Association*, 4(1):37–74.

Acemoglu, D., Egorov, G., and Sonin, K. (2008). Coalition Formation in Non-Democracies. *Review of Economic Studies*, pages 987–1009.

Acemoglu, D., Egorov, G., and Sonin, K. (2012). Dynamics and Stability of Constitutions, Coalitions, and Clubs. *American Economic Review*, 102(4):1446–1476.

Acemoglu, D. and Robinson, J. (2005). *Economic Origins of Dictatorship and Democracy*.

Acemoglu, D. and Robinson, J. (2012). *Why Nations Fail: Origins of Power, Poverty and Prosperity*.

Acemoglu, D., Robinson, J. A., and Verdier, T. (2004). Kleptocracy and divide-and-rule: A model of personal rule. *Journal of the European Economic Association*, 2(2-3):162–192.

Acemoglu, D., Ticchi, D., and Vindigni, A. (2010). Persistence of civil wars. *Journal of the European Economic Association*, 8(2-3):664–676.

Acemoglu, D. and Wolitzky, A. (2014). Cycles of conflict: An economic model. *American Economic Review*, 104(4):1350–1367.

Almas, I. and Johnsen, A. A. (2013). The cost of living in china: Implications for inequality and poverty. *Memorandum, Economics Department, University of Oslo*.

Auerbach, A. and Kotlikoff, L. (1987). *Dynamic Fiscal Policy*. Cambridge University Press, Cambridge.

Bai, C., Hsieh, C.-t., and Qian, Y. (2006a). The Return to Capital in China. *Brookings Papers on Economic Activity*, pages 61–88.

Bai, C.-E., Hsieh, C.-T., and Qian, Y. (2006b). The return to capital in china. *Brookings Papers on Economic Activity*, 37(2):61–102.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817.

Barbieri, K. (1996). Economic interdependence: A path to peace or a source of interstate conflict? *Journal of Peace Research*, 33(1):29–49.

Barr, N. and Diamond, P. (2008). *Reforming Pensions: Principles and Policy Choices*. Oxford University Press, Oxford.

Barro, R. J. and Lee, J.-W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104:184–198.

Besley, T. and Persson, T. (2009). The Origins of State Capacity: Property Rights, Taxation, and Politics. *American Economic Review*, 99(4):1218–1244.

Besley, T. and Persson, T. (2011a). Pillars of Prosperity.

Besley, T. and Persson, T. (2011b). *Pillars of prosperity: The political economics of development clusters*. Princeton University Press, Princeton.

Blattman, C. and Miguel, E. (2010). Civil war. *Journal of Economic Literature*, pages 3–57.

Bohn, H. (2011). Should public retirement plans be fully funded? *Journal of Pension Economics and Finance*, 10:195–219.

Brandt, L., Tombe, T., and Zhu, X. (2013). Factor market distortions across time, space and sectors in China. *Review of Economic Dynamics*, 16(1):39–58.

Brandt, L. and Zhu, X. (2000). Redistribution in a Decentralized Economy growth and Inflation in China under Reform. *Journal of Political Economy*, 108(2):422.

Brandt, L. and Zhu, X. (2010). Accounting for China's Growth. (4764).

Buiter, W. and Rahbari, E. (2011). Global growth generators: Moving beyond emerging markets and bric. *Citi Global Economics View, February*, 21.

Cai, F., Giles, J., and Meng, X. (2006). How well do children insure parents against low retirement income? an analysis using survey data from urban china. *Journal of Public Economics*, 90(12):2229–2255.

Caselli, F. and Coleman, W. J. (2013). On the theory of ethnic conflict. *Journal of the European Economic Association*, 11(1):161–192.

Chamon, M., Liu, K., and Prasad, E. S. (2013). Income uncertainty and household savings in china. *Journal of Development Economics*, 105:164–177.

Chan, K. W. and Buckingham, W. (2008). Is china abolishing the hukou system? *China Quarterly*, 195:582–606.

Chen, J. and Lu, C. (2011). Democratization and the Middle Class in China: The Middle Class's Attitudes toward Democracy. *Political Research Quarterly*, 64(3):705–719.

CIA (2014). The world factbook 2014. Central Intelligence Agency, Washington, DC.

Conesa, J. C. and Garriga, C. (2008). Optimal fiscal policy in the design of social security reforms. *International Economic Review*, 49(1):291–318.

Conesa, J. C. and Krueger, D. (2006). Social security reform with heterogeneous agents. *Review of Economic Dynamics*, 2(4):757–795.

Courdacier, N., Jin, K., and Gibaud, S. (2013). Fertility policies and social security reforms in china. Mimeo, London School of Economics.

Dirk, K. and Kubler, F. (2006). Pareto-improving social security reform when financial markets are incomplete. *American Economic Review*, 96(3):737–755.

Dollar, D. and Wei, S.-J. (2007). Das (wasted) kapital: Firm ownership and investment efficiency in china. Working Paper 13103, NBER.

Duan, C. and Sun, Y. (2006). Changes in the scope and definition of the floating population in china's censuses and surveys (in chinese). *Population Research*, 30(4):70–76.

Dunaway, S. V. and Arora, V. B. (2007). Pension reform in china: The need for a new approach. Working paper, IMF.

E, C. and JB., W. (2008). Old-age pension reform and modernization pathways: Lessons for china from latin america. *Journal of Aging Studies*, 22:74–87.

Eichengreen, B., Park, D., and Shin, K. (2013). Growth Slowdowns Redux: New Evidence on the Middle-income Trap.

Esteban, J., Mayoral, L., and Ray, D. (2012). Ethnicity and conflict: An empirical study. *The American Economic Review*, 102(4):1310–1342.

Esteban, J. and Ray, D. (2008). On the salience of ethnic conflict. *The American Economic Review*, 98(5):2185–2202.

Esteban, J. and Ray, D. (2011). Linking conflict to inequality and polarization. *The American Economic Review*, 101(4):1345–1374.

Fatás, A. and Mihov, I. (2009). Another Challenge to China's Growth. *Harvard Business Review*.

Fearon, J. D. and Laitin, D. D. (2000). Violence and the social construction of ethnic identity. *International Organization*, 54(4):845–877.

Fearon, J. D. and Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American political science review*, 97(1):75–90.

Feldstein, M. (1999). Social security pension reform in china. *China Economic Review*, 10(2):99–107.

Feldstein, M. and Liebman, J. (2006). Realizing the potential of china's social security pension system. *China Economic Times*, 24.

French, E. B. (2005). The effects of health, wealth, and wages on labour supply and retirement behaviour. *Review of Economic Studies*, 72(2):395–427.

Garriga, C. (1999). Optimal fiscal policy in overlapping generations models. Mimeo. University of Barcelona.

Ge, S. and Yang, D. T. (2012). Changes in China's Wage Structure. (6492).

Gill, I. and Kharas, H. (2007). *An East Asian Renaissance: Ideas for Economic Growth*.

Giordano, R., Depalo, D., Pereira, C., Eugène, B., Papapetrou, E., Perez, J. J., Reiss, L., Pereira, M. C., and Roter, M. (2011). The Public Sector Pay Gap In A Selection Of Euro Area Countries. *European Central Bank Working Papers*.

Goldsmith, B. E. (2013). International trade and the onset and escalation of interstate conflict: More to fight about, or more reasons not to fight? *Defence and Peace Economics*, 24(6):555–578.

Goodkind, D. M. (2004). China's missing children: The 2000 census underreporting surprise. *Population Studies*, 58(3):281–295.

Gourevitch, P. (1998). *We wish to inform you that tomorrow we will be killed with our families: Stories from Rwanda*. Macmillan, London.

Gowa, J. and Hicks, R. (2012). Conflict and commerce: New data about the great war.

Han, L. and Kung, J. K.-s. (2012). Fiscal Incentives and Policy Choices of Local Governments: Evidence from China.

He, H., Imrohoglu, S., and Sargent, T. J. (1997). Two computations to fund social security. *Macroeconomic Dynamics*, 1(1):7–44.

Horowitz, D. L. (2000). *Ethnic groups in conflict*. University of California Press, Berkeley.

Hsieh, C.-t. and Klenow, P. (2009a). Misallocation and Manufacturing TFP in China and India. *Quarterly Journal of Economics*, CXXIV(November).

Hsieh, C.-T. and Klenow, P. J. (2009b). Misallocation and manufacturing tfp in china and india. *Quarterly Journal of Economics*, 124(4):1403–1448.

Hsieh, C.-t. and Song, Z. (2013). Grasp the Large, Let Go of the Small: the Transformation of the State Sector in China.

Hu, Y. (2003). Quantitative analysis of the population transfers from rural areas to urban areas (in chinese). *Statistical Research*, 7:20–24.

Hu, Y.-W., Stewart, F., and Yermo, J. (2007). Pension fund investment and regulation: An international perspective and implications for china's pension system. Mimeo, OECD.

Huang, Y. (2008). *Capitalism with Chinese Characteristics: Entrepreneurship and the State*.

Hurd, M. D. (1989). Mortality risk and bequests. *Econometrica*, 57(4):779–813.

Islam, N., Dai, E., and Sakamoto, H. (2006). Role of tfp in china's growth. *Asian Economic Journal*, 20(2):127–159.

Jackson, M. O. and Morelli, M. (2007). Political bias and war. *The American Economic Review*, pages 1353–1373.

Johnson, D. G. (2003). Provincial migration in china in the 1990s. *China Economic Review*, 14(1):22–31.

Jr, D. F., J., R., and Weingast, B. R. (1997). Rationality of fear: political opportunism and ethnic conflict. *Military Intervention in Civil Wars*.

Kaplan, G. and Schulhofer-Wohl, S. (2012). Understanding the long-run decline in interstate migration. Working Paper 697, Federal Reserve Bank of Minneapolis.

Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature*, 49(4):961–1075.

Lavely, W. (2001). First impressions from the 2000 census of china. *Population and Development Review*, 27(4):755–769.

Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *Manchester School*, pages 139–91.

Liang, Z. and Ma, Z. (2004). China's floating population: New evidence from the 2000 census. *Population and Development Review*, 30(3):467–488.

Lu, J. (2011). Impacts of demographic transition on future economic growth: China's case study. *Paper presented at the conference China and and the West*, pages 1950–2050.

Martin, P., Mayer, T., and Thoenig, M. (2008). Make trade not war? *The Review of Economic Studies*, 75(3):865–900.

Melly, B. (2002). Public-private Sector Wage Differentials in Germany: Evidence from Quantile Regression. pages 1–26.

Mueller, R. E. (1998). Public-private Sector Wage Differentials in Canada: Evidence from Quantile Regressions. *Economics Letters*, 60(2):229–235.

Musacchio, A. and Lazzarini, S. G. (2012). Leviathan in Business : Varieties of State Capitalism and their Implications for Economic Performance. pages 1–63.

Newbury, M. C. (1978). Ethnicity in rwanda: the case of kinyaga. *Africa: Journal of the International African Institute*, 48(1):17–29.

Newbury, M. C. (1988). *The Cohesion of Oppression: clientship and ethnicity in Rwanda, 1860-1960*. Columbia University Press, New York.

Newbury, M. C. (1998). Ethnicity and the politics of history in rwanda. *Africa Today*, pages 7–24.

Nishiyama, S. and Smetters, K. (2007). Does social security privatization produce efficiency gains? *Quarterly Journal of Economics*, 122(4):1677–1719.

Nordhaus, W. (2007). Critical assumptions in the stern review on climate change. *Science*, 317(5835):201–202.

Padro-i Miquel, G. (2007). The control of politicians in divided societies: the politics of fear. *The Review of Economic Studies*, 74(4):1259–1274.

Park, A., Shen, Y., Strauss, J., and Zhao, Y. (2012). Relying on whom? poverty and consumption financing of china's elderly. In James, S. and Majmundar, M., editors, *Aging in Asia Findings From New and Emerging Data Initiatives*. National Academies Press, Washington (DC).

Posner, E. A., Spier, K. E., and Vermeule, A. (2010). Divide and conquer. *Journal of Legal Analysis*, 2(2):417–471.

Poterba, J. M. and Rueben, K. S. (1994). The Distribution of Public Sector Wage Premia: New Evidence Using Quantile Regression Methods. *NBER Working Paper*, (May).

Pottier, J. (2002). *Re-imaging Rwanda*. Cambridge University Press, Cambridge.

Prunier, G. (1995). *The Rwanda crisis: History of a genocide*. Columbia University Press, New York.

Prunier, G. (2008). *Africa's World War: Congo, the Rwandan genocide, and the making of a continental catastrophe*. Oxford University Press, Oxford.

Reinikka, R. and Svensson, J. (2004). Local Capture: Evidence from a Central Government Transfer Program in Uganda. *Quarterly Journal of Economics*, (May).

Robinson, J. A. and Torvik, R. (2004). White Elephants. *Journal of Public Economics*, 89:197–210.

Rogerson, R. and Wallenius, J. (2009). Retirement in a life cycle model of labor supply with home production. Working paper, University of Michigan.

Rohner, D. (2011). Reputation, group structure and social tensions. *Journal of Development Economics*, 96(2):188–199.

Rohner, D., Thoenig, M., and Zilibotti, F. (2013). War signals: A theory of trade, trust, and conflict. *The Review of Economic Studies*, 80(3):1114–1147.

Salditt, F., Whiteford, P., and Adema, W. (2007). Pension reform in china: Progress and prospects. Social, Employment and Migration Working Paper 53, OECD.

Sin, Y. (2005). China: Pension liabilities and reform options for old age insurance. Working Paper 2005–1, World Bank.

Song, Z., Storesletten, K., and Zilibotti, F. (2011a). Growing Like China. *American Economic Review*, 101(February):196–233.

Song, Z., Storesletten, K., and Zilibotti, F. (2011b). Growing like china. *American Economic Review*, 101(1):196–233.

Song, Z., Storesletten, K., and Zilibotti, F. (2014). Growing (with capital controls) like china. *IMF Economic Review (forthcoming)*.

Song, Z. and Yang, D. T. (2010). Life cycle earnings and the household saving puzzle in a fast-growing economy. Mimeo, Chinese University of Hong Kong.

Storesletten, K. (2000). Sustaining fiscal policy through immigration. *Journal of Political Economy*, 108(2):300–323.

Storesletten, K. and Zilibotti, F. (2014). China's Great Convergence and Beyond. *Annual Economic Review*.

Suqing, G. and Yang, D. (2014). Changes in china's wage structure. *Journal of the European Economic Association*, 12(2):300–336.

Tabellini, G. (2008). The scope of cooperation: Values and incentives. *The Quarterly Journal of Economics*, 123(3):905–950.

The Economist (2009). Special report: The new middle classes in emerging markets.

Tsai, K. (2007). *Capitalism without Democracy: The Private Sector in Contemporary China*.

Woodward, S. L. (1995). *Balkan tragedy: Chaos and dissolution after the Cold War*. Brookings Institution Press, Washinton, D.C.

Xia, Q., Song, L., Li, S., and Appleton, S. (2011). Effect of Public Sector and SOEs on Wage Inequality of Urban China. pages 1988–2007.

Yang, J. (2011). Population change and poverty among the elderly in transitional china. Mimeo, Center for Population and Development Studies, Renmin University of China.

Yang, J. and Chen, Z. (2010). Economic poverty among the elderly in the era of family change: A quantitative and qualitative analysis. *Population Research*, 34(5):51–67.

Yearbook, C. S. *various issues*. National Bureau of Statistics of China.

Zeng, Y. (2007). Options for fertility policy transition in china. *Population and Development Review*, 33(2):215–246.

Zhai, Z. and Wei, C. (2007). Chinese fertility in the 1990s (in chinese). *Population Research*, 31(1):19–32.

Zhang, G. and Zhao, Z. (2006). Reexamining china's fertility puzzle: Data collection and quality over the last two decades. *Population and Development Review*, 32(2):293–321.

Zhang, J., Zhao, Y., Park, A., and Song, X. (2005). Economic returns to schooling in urban china, 1988 to 2001. *Journal of Comparative Economics*, 33(4):730–752.

Zhang, W. and Cui, H. (2003). Estimation of accuracy of 2000 national population census data (in chinese). *Population Research*, 27(4):25–35.

Zhao, Y. and Xu., J. (2002). China's urban pension system: Reforms and problems. *The Cato Journal*, 21.3:395–414.

Zhou, Y. and Ma, L. J. C. (2003). China's urbanization levels: Reconstructing a baseline from the fifth population census. *China Quarterly*, 173:176–196.

Zimmerman, W. (1995). Last ambassador-a memoir of the collapse of yugoslavia. *Foreign Affairs*, 74(2).